STAT 231 — Spring 2024: Class Notes

Jiucheng Zang

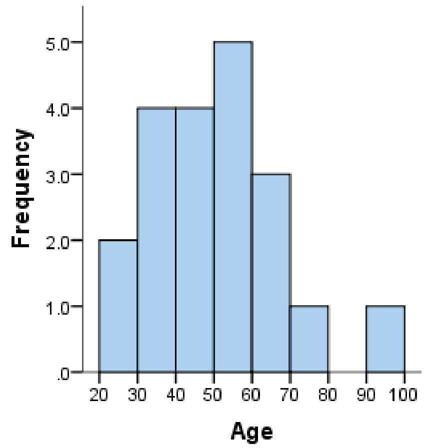
May 13th 2024

Contents

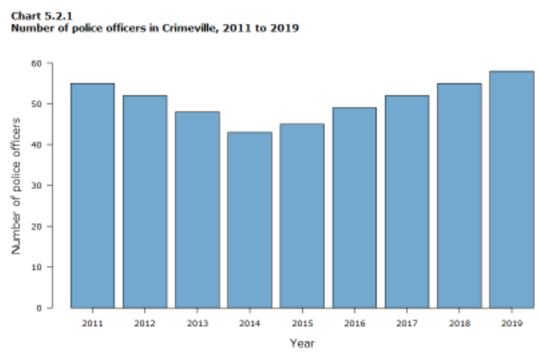
I. Data Summaries

i. Different Graphs

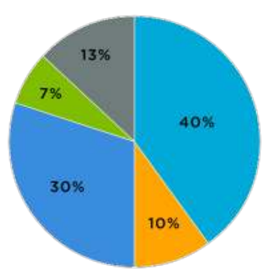
- **Histogram:** A bar graph that shows the frequency of data within a certain range.



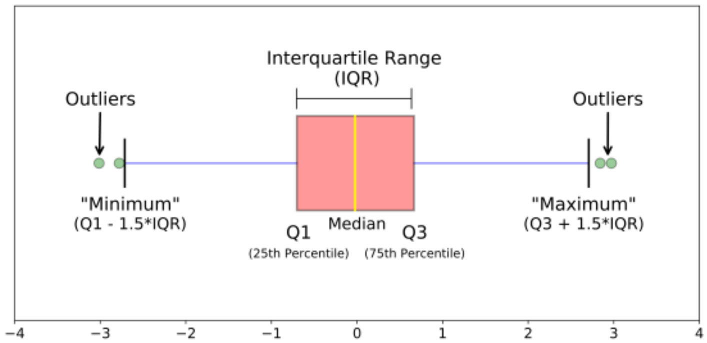
- **Bar Chart:** A graph same as histogram but for x-axis with no meaningful quantitative order.



- **Pie Chart:** A circle graph that shows the proportion of data in each category.



- **Box Plot:** A graph that shows the distribution of data based on the five-number summary.



Shape of dataset;

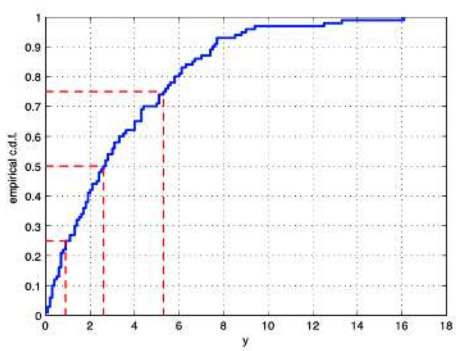
Skewness looking at Box plot.

Smallest non-outlier 'whiskers' are the 1st and 3rd quartiles.

ii. Statistical Characteristic

Empirical c.d.f.: is a step function that jumps up by $1/n$ at each data point.

$$F(y) = \frac{\text{Number of data points} \leq y}{\text{Total number of data points}}$$



Sample Correction: $n-1$ instead of n in the denominator of the sample variance.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

IMPORTANT NOTES:

1. Strong linear relationship doesn't imply causation.
2. Correlation value doesn't imply causation.

iii. Statistical Inference

1. Descriptive Statistics → Graph / Compute Summary Statistics.
2. Spatial Inference → Inductive reasoning / Deductive reasoning.
   (a) Estimation problems.
   (b) Hypothesis testing problems.
   (c) Prediction problems.

II. Statistical models and maximum likelihood estimation

Gaussian distribution (μ, δ) or Normal distribution (μ, δ^2):

> "How well does the model fit the data?"

i. General Family of models, $P(Y = y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$

Estimation of unknown parameter θ:

- "μ" is the 'true' value of θ.
- $\hat{\theta}$ is the estimate of θ.

Sequence chooses a model:

1. Collect and examine the data.
2. Propose a model.
3. Fit the model.
4. Check the model.
5. Return to step 2 or exit the loop.
6. Draw conclusions.

Table 2.1
Properties of discrete versus continuous random variables

Property	Discrete	Continuous
cumulative distribution function	$F(x) = P(X \le x) = \sum_{t \le x} P(X = t)$ F is a right continuous step function for all $x \in \Re$	$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ F is a continuous function for all $x \in \Re$
probability (density) function	$f(x) = P(X = x)$	$f(x) = \frac{d}{dx}F(x) \neq P(X = x) = 0$
Probability of an event	$P(X \in A) = \sum_{x \in A} P(X = x)$ $= \sum_{x \in A} f(x)$	$P(a < X \le b) = F(b) - F(a)$ $= \int_{a}^{b} f(x)\,dx$
Total probability	$\sum_{all\ x} P(X = x) = \sum_{all\ x} f(x) = 1$	$\int_{-\infty}^{\infty} f(x)\,dx = 1$
Expectation	$E[g(X)] = \sum_{all\ x} g(x) f(x)$	$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

ii. Different distributions

1. Binomial Distribution

The discrete random variable Y has a binomial distribution if its probability function is of the form

$$P(Y = t; \theta) = f(y; \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}$$

where $0 \le \theta \le 1$ and $y = 0, 1, 2, ..., n$.

Where θ is a $0 < \theta < 1$. For convenience we write $Y \sim B(n, \theta)$. Recall that $E(Y) = n\theta$ and $Var(Y) = n\theta(1 - \theta)$.

2. Poisson Distribution

The discrete random variable Y has a Poisson distribution if its probability function is of the form

$$P(y, \theta) = \frac{\theta^y e^{-\theta}}{y!}$$

where $\theta > 0$ and $y = 0, 1, 2, ...$

Where θ is a $\theta > 0$. For convenience, we write $Y \sim P(\theta)$. Recall that $E(Y) = \theta$ and $Var(Y) = \theta$.

3. Exponential Distribution

The continuous random variable Y has an Exponential distribution if its probability density function is of the form

$$f(y;\theta) = \frac{1}{\theta}e^{-y/\theta}$$

where $\theta > 0$ and $y > 0$.

Where θ is a $\theta > 0$. For convenience, we write $Y \sim Exp(\theta)$. Recall that $E(Y) = \theta$ and $Var(Y) = \theta^2$.

4. Gaussian (Normal) Distribution

The continuous random variable Y has a Gaussian or Normal distribution if its probability density function is of the form

$$f(y;\mu,\delta) = \frac{1}{\sqrt{2\pi}\delta}\exp\left[-(y-\mu)^2/2\delta^2\right]$$

where $-\infty < \mu < \infty$ and $\delta > 0$.

Where μ is a $-\infty < \mu < \infty$ and δ is a $\delta > 0$. For convenience, we write $Y \sim N(\mu, \delta^2)$. Recall that $E(Y) = \mu$ and $Var(Y) = \delta^2$.

Note that in the former case, $G(\mu, \delta)$, the second parameter is the standard deviation δ.

5. Multinomial distribution

The Multinomial distribution is a multivariate distribution in which the discrete random variable's $Y_1, Y_2, ..., Y_k (k \geq 2)$ have the joint probability functtion

$$P(y_1, y_2, ..., y_k; \theta_1, \theta_2, ..., \theta_k) = \frac{n!}{y_1!y_2!...y_k!}\theta_1^{y_1}\theta_2^{y_2}...\theta_k^{y_k}$$

where $\theta_1, \theta_2, ..., \theta_k$ are the parameters of the distribution and $y_1, y_2, ..., y_k$ are non-negative integers such that $y_1 + y_2 + ... + y_k = n$. The parameters $\theta_1, \theta_2, ..., \theta_k$ are subject to the constraints $0 \leq \theta_i \leq 1$.

iii. Point Estimates and Maximum Likelihood Estimation

Def :
A point estimate of a parameter is a value of function of the observed data $y_1, y_2, ..., y_n$ and other known quantities that is used as an estimate of the parameter. We use the notation $\hat{\theta}$ to denote a point estimate of a parameter θ.

	Parameter θ	Estimator $\tilde{\theta}$	Estimate $\hat{\theta}$
$G(\mu, \delta)$	μ (population)	$\tilde{\mu} = \frac{1}{n}\sum Y_i$	$\hat{\mu} = \frac{1}{n}\sum y_i$
$Bin(n, \theta)$	θ (population proportion)	$\tilde{\theta} = \frac{Y}{n}$	$\hat{\theta} = \frac{y}{n}$

iv. Likelihood Function

$$f(y) \text{ V.S } f(y;\theta)$$

$Y \sim Bin(25, \theta)$

Now we have experience with tossing a coin 25 times, and we want to estimate the probability of getting a head. From the data, we have 10 heads. We can use the following formula to estimate the probability of getting a head:

$P(Y = y; \theta) = \binom{25}{y}\theta^y(1-\theta)^{25-y}, 0 < \theta < 1, y = 0, 1, 2, \ldots, 25$

Maximizes expression, so we need to find the maximum likelihood estimator. (By take the derivative and set it to 0)

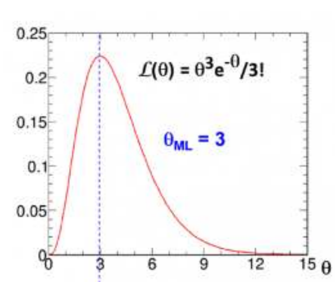
$$\frac{dP(Y = y;\theta)}{d\theta} = \binom{25}{10}\theta^9(1-\theta)^{14}(10 - 25\theta) = 0$$
$$\theta = \frac{10}{25} = 0.4$$

Data most likely to be generated by a coin with a probability of 0.4.

Method of Maximum Likelihood (m.l.e.)

Likelihood function for θ (discrete random variable):

$$L(\theta) = L(\theta; y) = P(Y = y; \theta) \text{ for } \theta \in \Omega$$



Binomial Data,

m.l.e. of θ is $\hat{\theta} = \frac{y}{n}$

Import: It's possible for any $\hat{\theta}$ but $\hat{\theta} = 0.4$ is the most likely.

Note: $\binom{n}{y}$ doesn't affect what value of θ, we are only maximizes $L(\theta)$.

$$\theta^y(1-\theta)^{n-y}$$

$P(Y = 10; \theta = 0.4) = \binom{25}{10}0.4^{10}0.6^{15} = 0.161$
$P(Y = 10; \theta = 0.25) = \binom{25}{10}0.25^{10}0.75^{15} = 0.042$

$$L(\theta_1)/L(\theta_2)$$

The data are approx 4 times more likely to be generated by a coin with a probability of 0.4 than a coin with a probability of 0.25.

v. Relative Likelihood Function

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \text{ for } \theta \in \Omega$$

$$R(\theta) = \frac{P(Y = y; \theta)}{P(Y = y; \hat{\theta})} = \frac{\theta^y(1-\theta)^{n-y}}{\hat{\theta}^y(1-\hat{\theta})^{n-y}} \text{ where } \hat{\theta} = \frac{y}{n}$$

It's a ratio of the likelihood of the data being generated by a coin with a probability of θ to the likelihood of the data being generated by a coin with a probability of $\hat{\theta}$.

$$R(\hat{\theta}) = 1$$

vi. Log Likelihood Function

$$\ell(\theta) = \log L(\theta), \theta \in \Omega$$

$L(\theta)$ and $\ell(\theta)$ are maximized for the same value of θ.

Write $L(\theta) \; \alpha \; \theta^y(1-\theta)^{n-y}$

$$\ell(\theta) = 10\ln(\theta) + 15\ln(1-\theta)$$

$$
\begin{aligned}
\ell'(\theta) = \frac{d\ell(\theta)}{d\theta} = \frac{10}{\theta} - \frac{15}{1-\theta} = 0 \\
\frac{10}{\hat{\theta}} = \frac{15}{1-\hat{\theta}} \\
\hat{\theta} = \frac{10}{25} = 0.4
\end{aligned}
$$

vii. Multiparameter Likelihood Function

$Y_1, Y_2, \ldots, Y_n$ are random sample. Assume to be independent and identically distributed (i.i.d) from a population with a probability distribution $P(Y = y; \theta)$, where θ is an unknown parameter.

$$
\begin{aligned}
L(\theta) &= P(\text{ observed the data } y_1, y_2, \ldots, y_n; \theta) \\
&= P(Y_1 = y_1; \theta) \times P(Y_2 = y_2; \theta) \times \cdots \times P(Y_n = y_n; \theta) \\
&= \prod_{i=1}^{n} P(Y_i = y_i; \theta)
\end{aligned}
$$

Geometric model θ : probability of success

1. $Y = \#$ rolls before "S" $Y \sim Geo(\theta)$

$$
P(Y = y; \theta) = (1-\theta)^y \theta
$$

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} P(Y_i = y_i; \theta) \\
&= \prod_{i=1}^{n} (1-\theta)^{y_i} \theta \\
&= (1-\theta)^{\sum_{i=1}^{n} y_i} \theta^n \\
&= (1-\theta)^{n\overline{y}} \theta^n
\end{aligned}
$$

$\ell(\theta) = n\overline{y} \ln(1-\theta) + n \ln(\theta)$
$\ell'(\theta) = -\frac{n\overline{y}}{1-\theta} + \frac{n}{\theta}$
$\ell'(\theta) = 0 : \frac{n}{\hat{\theta}} = \frac{n\overline{y}}{1-\hat{\theta}} \Leftrightarrow \hat{\theta} = \frac{1}{\overline{y}+1}$

Example:

$\bar{y} = \frac{5+3+\cdots+7}{10} = 5.4$
$\therefore \hat{\theta} = \frac{1}{5.4+1} \approx 0.156$

Possion model

$Y \approx Poi(\theta), P(Y = y; \theta) = \begin{cases} \frac{e^y e^{-\theta}}{y!}, & y = 0, 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$

$$L(\theta) = \prod_{i=1}^{n} P(Y_i = y_i; \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i} e^{-\theta}}{y_i!} = \dots$$

$$\hat{\theta} = \bar{y}$$

Two independent dataset Y_1 and Y_2, their likelihood function can be product.

$$L(\theta) = L_1(\theta) \times L_2(\theta) \theta \in \Omega$$

where $L_i(\theta) = P(Y_i = y_i; \theta)$

viii. Likelihood function for continuous random variable

$$P(a \leq Y \leq b; \theta) = \int_a^b f(y; \theta) dy \text{ probability density function}$$

$Y \sim Exp(\theta),\ f(y; \theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, y \geq 0, \theta > 0$

$L(\theta) = \prod_{i=1}^{n} f(y_i; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{y_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^{n} y_i}{\theta}} = \frac{1}{\theta^n} e^{-\frac{n\bar{y}}{\theta}}$

$\ell(\theta) = -n \ln(\theta) - \frac{n\bar{y}}{\theta}$

$\ell'(\theta) = -\frac{n}{\theta} + \frac{n\bar{y}}{\theta^2} = 0 \Leftrightarrow \hat{\theta} = \bar{y}$

ix. Find maximum likelihood function for Gaussian distribution

$Y \sim G(\mu, \delta)$,

$$L(\theta) = \prod_{i=1}^{n} f(y_i; \mu, \delta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(y_i - \mu)^2}{2\delta^2}} = \frac{1}{(2\pi\delta^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2\delta^2}}$$

$$L(\theta) = -n \ln(\theta) - \frac{1}{2\delta^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

$$\ell(\mu, \delta) = -\frac{n}{2} \ln(2\pi\delta^2) - \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2\delta^2}$$

$$\ell'(\mu, \delta) = \begin{cases} \frac{\sum_{i=1}^{n}(y_i - \mu)}{\delta^2} = 0 & \text{for } \mu \\ -\frac{n}{\delta} + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{\delta^3} = 0 & \text{for } \delta \end{cases}$$

$\hat{\mu} = \bar{y},\ \hat{\delta} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}}$

We estimate μ and δ by $\bar{y}$ and $\hat{\delta}$.

x. Invariance Property of Maximum Likelihood Estimate

Example:

> Suppose someone bet \$10 that the Canadians would score at least 3 goals in their next game. If they score 2 goals or fewer you win \$10, if they score 3 goals or more you lost \$10.
>
> Should you take the bet?

Invariance Property of Maximum Likelihood Estimate of θ

$$P(Y = 3) = \frac{e^{\theta} \theta^3}{3!}$$

If $\hat{\theta}$ is the maximum likelihood estimate of θ, then $g(\hat{\theta})$ is the maximum likelihood estimate of $g(\theta)$
We want to know the probability of Canadiens will score at least 3 goals.

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) \\ &= 1 - \sum_{y=0}^{2} P(Y = y) \\ &= 1 - \sum_{y=0}^{2} \frac{\theta^y e^{-\theta}}{y!} \end{aligned}$$

The maximum likelihood estimate of $P(Y \geq 3)$ is:

$$1 - \sum_{y=0}^{2} \frac{e^{-\hat{\theta}}\hat{\theta}y}{y!} = 1 - \sum_{y=0}^{2} \frac{e^{-3}3^y}{y!} = 0.577$$

xi. Checking Model Fit

In order to study a dataset we typically assume a model in which $Y = (Y_1, Y_2, \ldots, Y_n$ is a potential random sample from a distribution which is a member of the family of models.

$$f(y;\theta), \theta \in \Omega$$

- Compare observed frequencies with expected frequencies calculate using the assumed model (discrete and continuous distributions.)
- Examine a Normal or Gaussian qqplot.

We'll start by comparing observed and expected frequencies: if the model is suitable, these should be 'close'.

We can compare the observed frequencies based on the data, with the expected frequencies calculated using probabilities based on the Poisson(θ) model.

Example:

Alpha-Paticles

Number of Alpha-Particles Detected	(Observed) Frequency f_j	Expected Frequency e_j
0	57	?
1	203	?
2	383	?
3	525	?
4	532	?
5	408	?
6	273	?
7	139	?
8	45	?
9	27	?
10	10	?
11	6	?
Total	2608	2608

If we let Y present the number of alpha-particles observed in a fixed period of time, then $Y \sim Poisson(\theta)$.
θ = mean # of alpha-particles observed for a fixed time period.

$$\bar{y} = \frac{1}{2608}[57(0) + 204(1) + \cdots + 6(11)] \approx 3.87$$

If $Y \sim Poisson(\theta)$, the probability of observing $Y = y$ for a particular experiment is:

$$P(Y = y) = \frac{e^{-\theta}\theta^{y}}{y!}$$

We've estimated $\hat{\theta} = \bar{y} = 3.87$, and so if our Poisson model is correct the probability of observing exactly y alpha-particles in a fixed time period is:

$$P(Y = y) = \frac{e^{-3.87}3.87^{y}}{y!}, \quad y = 0, 1, 2, \ldots, 11$$

So, e.g., if our Poisson model is correct the probability of observing no alpha-particles in a fixed interval is $P(Y = 0) = \frac{e^{-3.87}3.87^{0}}{0!} = e^{-3.87} \approx 0.021$.
Since $P(Y = 0) = 0.021$, we would expect to observe $0.021 \times 2608 \approx 54.7$ experiments with no alpha-particles.

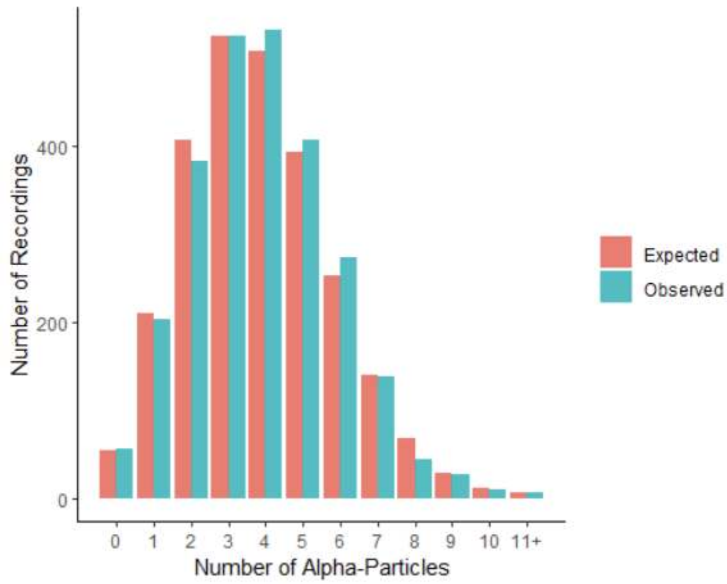
Number of Alpha-Particles Detected	Frequency f_j	Expected Frequency e_j
0	57	? $e_0 = 54.42$
1	203	?

More generally, we would expect to observe

$$e_j = (2608)P(Y = j) = (2608)\frac{e^{-3.87}3.87^{j}}{j!}, \quad j = 0, 1, 2, \ldots, 11$$

The Poisson distribution is discrete but goes off to infinity, so we can't calculate all the expected frequencies.

$$\begin{aligned} e_{11+} &= e_{11} + e_{12} + e_{13} + \ldots \\ &= (2608)P(Y = 11) + (2608)P(Y = 12) + \ldots \\ &= (2608)P(Y \geq 11) \\ &= (2608)[1 - P(Y \leq 10)] \\ &= (2608)[1 - \sum_{j=0}^{10} P(Y = j)] \\ &= 2608 - \sum_{j=0}^{10} e_j \end{aligned}$$



xii. Normal Distribution

This is also reasonable for the Normal distribution $Y \sim G(\mu, \sigma)$.
We need to estimate the values of the unknown parameters μ and σ based on the data $y_1, y_2, \ldots, y_n$.
We will estimate μ by $\bar{y}$ and σ using the sample mean (the mle) and sample standard deviation (not the mle) reepectively:

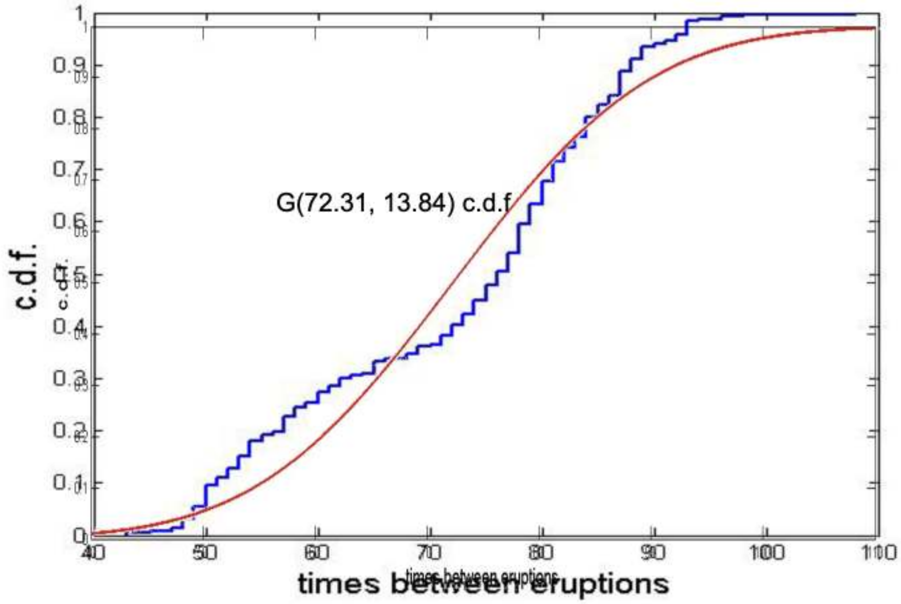
$$\hat{\mu} = \bar{y} = 159.77, \quad s^2 = 36.36, \quad \hat{\sigma} = s = 6.03$$

For example, if $Y \sim G(159.77, 6.03)$, then the probability a randomly chosen elderly woman has a height between 155 and 156 cm is:

$$\begin{aligned} P(155 \leq Y \leq 156) &= P\left(\frac{155 - 159.77}{6.03} \leq \frac{Y - 159.77}{6.03} (Z \leq \frac{156 - 159.77}{6.03}\right) \\ &= P(-0.79 \leq Z \leq -0.63) \\ &= P(Z \leq -0.63) - P(Z \leq -0.79) \\ &= (1 - P(Z \leq 0.63)) - (1 - P(Z \leq 0.79)) \\ &= P(Z \leq 0.79) - P(Z \leq 0.63) \\ &= 0.78524 - 0.73565 \\ &= 0.04959 (Z \sim N(0, 1)) \end{aligned}$$

$$\begin{aligned} p_j &= P\left(Y \in [a_{j-1}, a_j)\right) \\ &= P\left(a_{j-1} \leq Y \leq a_j\right) \\ &= P\left(\frac{a_{j-1} - 159.77}{6.03} \leq Z < \frac{a_j - 159.77}{6.03}\right) \text{ where } Z \sim G(0, 1) \end{aligned}$$

xiii. Normal cdf v.s. Empirical cdf



1. Normal QQ Plots

To compare an empirical c.d.f. with a Normal or Gaussian c.d.f. more easily we can use a graphical technique called a qqplot. The basic idea is that we want to create a plot based on the data, which will be approximately a straight line if a Normal model fits the data well.

If we want to check if a $G(\mu, \sigma)$ model fits the data well, we could compare the sample quantile $q(p)$ to the theoretical quantiles $Q(p)$ from the $G(\mu, \sigma)$ distribution.

$Q(p)$ satisfies $P(Y \leq Q(p)) = p$ for $Y \sim G(\mu, \sigma)$.

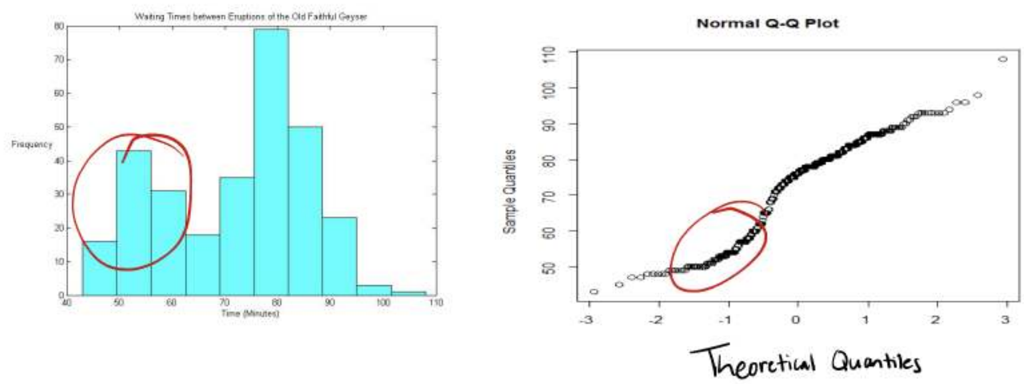
Let $Q_z(p)$ be the pth quantile for a $G(0,1)$. Since $\frac{Y-\mu}{\sigma} \sim G(0,1)$, $Q(p) = \mu + \sigma Q_z(p)$.

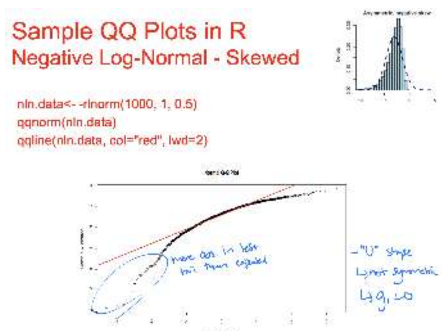
$$P(Y \leq Q(p)) = p \Rightarrow P(Z \leq \frac{Q(p)-\mu}{\sigma}) = p$$

If we plot the points $\left(Q_z\left(\frac{i}{n+1}\right), q\left(\frac{i}{n+1}\right)\right)$, we should see a straight line if the data is normally distributed.

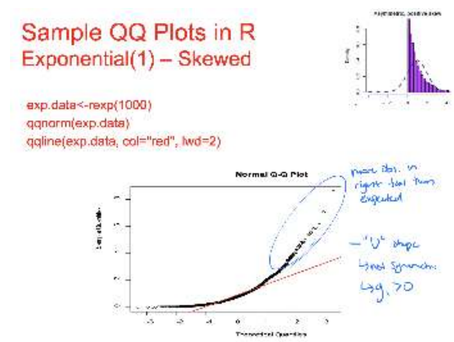
It looks like there is a problem (lack of agreement) on the left side. Does this make sense?

A Normal model doesn't appear to fit this data very well.

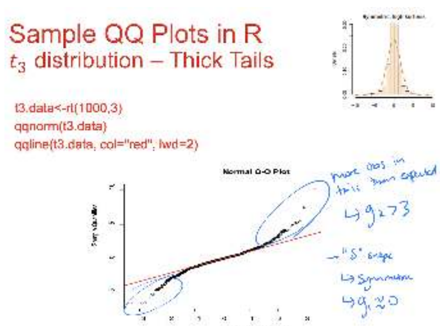




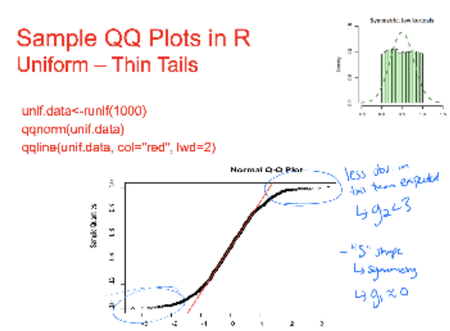
(a)



(b)



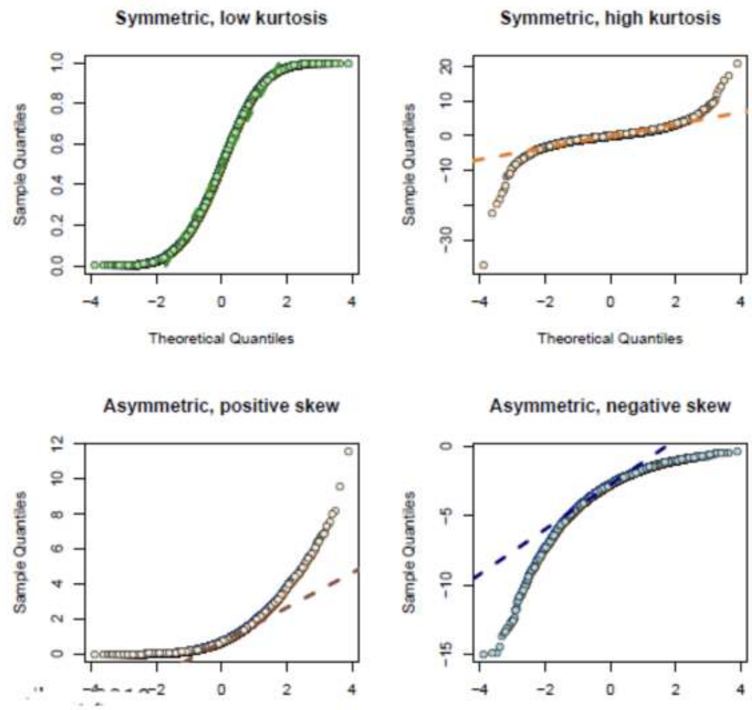
(c)



(d)

Figure 1: qqplot for different show out of normal distribution

Key Things to Look for in QQ Plots



III. Planning and Conducting Empirical Studies

Organizing a Statistical Problem:

PPDAC

- **P**roblem: Develop a clear statement of what we are trying to learn.
- **P**lan: Determine how we will carry out the empirical study.
- **D**ata: Collect the data according to the plan.
- **A**nalysis: Analyze the data to answer the questions posed in the problem step.
- **C**onclusion: Draw conclusions about what has been learned.

i. Normal Approximations

- Binomial: $\frac{y}{n} \sim G\left(\theta, \sqrt{\frac{\theta(1-\theta)}{n}}\right)$ approximately
- Poisson: $\overline{Y} \sim G\left(\theta, \sqrt{\frac{\theta}{n}}\right)$ approximately
- Exponential: $\overline{Y} \sim G\left(\theta, \frac{\theta}{\sqrt{n}}\right)$ approximately

Key concept: No matter the distribution of the population, the CLT suggests that for sufficiently large samples our sampling distribution is approximately Gaussian.

ii. Interval Estimation

We discussed the idea of point estimates and point estimators and the differences between them.

An estimate will pro us with a single value, or ”point”.

To indicate the uncertainty in an estimate we will use an **interval estimate** on the form

$$[L(y), U(y)]$$

where $L(y)$ and $U(y)$ are both functions of the observed data $y = (y_1, y_2, \ldots, y_n)$.

iii. Confidence Intervals and Pivotal Quantities

Recall: In general, an interval estimate for θ based on observed data, y, takes on the form $[L(y), U(y)]$.

$$\begin{aligned} p &= P(\theta \in [L(y), U(y)]) \\ &= P(\theta \text{ is contained in the random interval } [L(Y), U(Y)]) \end{aligned}$$

θ is the true (but UNKNOWN) value of the parameter.

1. Coverage Probability

Def :
The value of

$$P(\theta \in [L(Y), U(Y)]) = P(L(Y) \leq \theta \leq U(Y))$$

is the **coverage probability** for the interval estimator $[L(Y), U(Y)]$.

We try to choose interval estimators (rules)

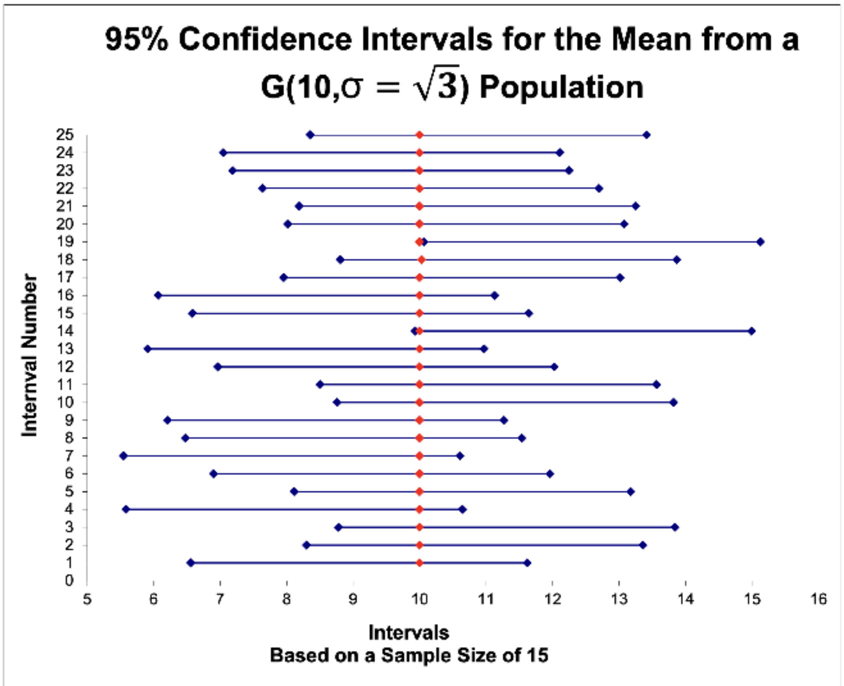
- The coverage probability is large (value close to 1)
- The interval is as narrow as possible

2. Confidence Intervals

Def :
A 100p% confidence interval for a parameter is an interval estimate $[L(Y), U(Y)]$ for which

$$P(\theta \in [L(Y), U(Y)]) = P(L(Y) \leq \theta \leq U(Y)) = p$$

The value p is called the confidence coefficient for the confidence interval.



25 intervals were constructed from a Gaussian population with mean 10 and standard deviation of 3.

24 of the 0.95 confidence intervals contained the true mean of 10.

This is in-line with what we would have expected.

In this case, if we were to repeatedly sample from the same population many, many times, then we would expect 95% of the intervals to contain the true mean, and 5% of the intervals to not contain the true mean.

3. Useful Resources

For more information, you can visit our website at `https://shiny.math.uwaterloo.ca/sas/stat231/coverage/`.

iv. Pivotal Quantities

A pivotal quantity $Q = g(Y;\theta)$ is a function of the data Y and the unknown parameter θ such that the distribution of the random variable is fully known.

Probability statements such as $P(Q \leq a)$ and $P(Q \geq b)$ depend on a and b but not θ or any other unknown parameter.

Using pivotal quantities, we can construct confidence intervals and conduct hypothesis tests.

Since $g(Y;\theta)$ is a pivotal quantity whose distribution is completely known, we can determine a and b such that

$$P(a \le g(Y;\theta) \le b) = P[L(Y) \le \theta \le U(Y)] = p$$

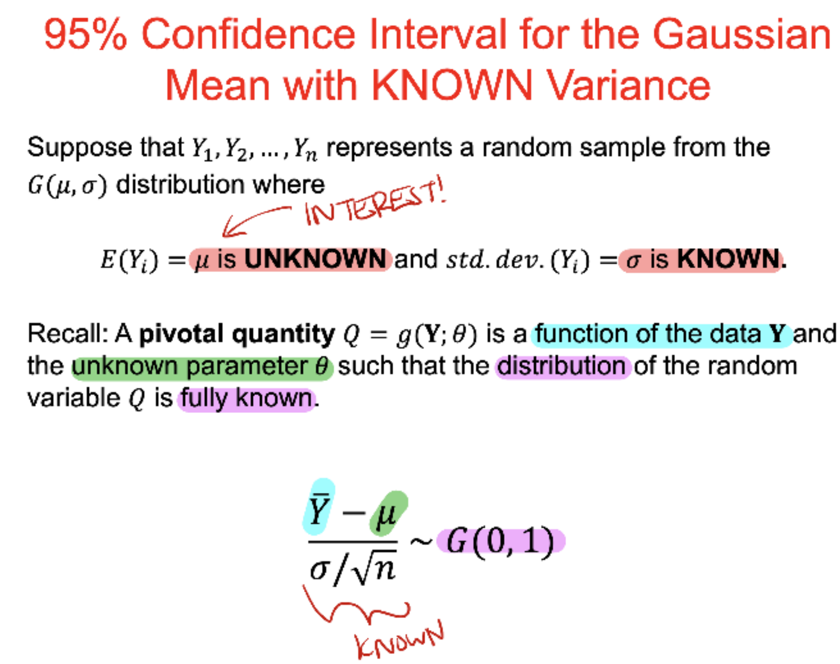
Without knowing θ. For the observed data y, we determine the interval $[L(y), U(y)]$ and this represents a $100p\%$ confidence interval for θ.

Example:

Suppose that $Y_1, Y_2, \ldots, Y_n$ represents a random sample from the $G(\mu, 10)$ distribution where $E(Y_i) = \mu$ is UNKNOWN and std. Deviation is known to be 10. $std.dev.(Y_i) = \sigma = 10$ (KNOWN)

What is the distribution of $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$?

Gaussian($\mu = 0, \sigma = 1$) (From the Central Limit Theorem)



We can use this pivotal quantity to build a 95% confidence interval for the population mean, μ, using

$$p = 0.95 = P(a \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le b)$$

$$\begin{aligned}
0.95 &= P(a \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le b) \\
&= P(-z \le Z \le z), Z \sim G(0,1) \\
&= 2P(Z \le z) - 1 \\
&\Rightarrow P(Z \le z) = 0.975
\end{aligned}$$

From the Z-table, we find that $z = 1.96$. Therefore, the 95% confidence interval for μ is

$$\begin{aligned}
0.95 &= P(-1.96 \leq \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq 1.96) \\
&= P(\underbrace{\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}}}_{L(Y)} \leq \mu \leq \underbrace{\bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}}_{U(Y)}) \\
&= P(\mu \in [\bar{Y} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{Y} + 1.96\frac{\sigma}{\sqrt{n}}]) \\
&= P(\mu \in [L(Y), U(Y)])
\end{aligned}$$

In general,

$$\left[\bar{Y} - c\frac{\sigma}{\sqrt{n}}, \bar{Y} + c\frac{\sigma}{\sqrt{n}}\right] = [L(y), U(y)]$$

Represents a confidence interval for μ based on the data, $y = (y_1, y_2, \dots, y_n)$.

- The above interval is symmetric about the sample mean, $\bar{y}$.
- It is also the narrowest possible interval.
- The interval becomes larger/wider as the confidence level increases.
- The interval becomes smaller/narrower as the sample size increases.(As confidence increases, the sample size increases square times)
- The interval becomes smaller/narrower as the standard deviation(σ) decreases.

Componets of a CI for μ Gaussian Population with δ known:

$$\left[\bar{y} - c\frac{\sigma}{\sqrt{n}}, \bar{y} + c\frac{\sigma}{\sqrt{n}}\right]$$

Suppose $\bar{y} = 10, \sigma = 2, n = 10$
Calculate CIs for μ where

1. 100p% is 90%, 95%, 98%, keeping everything else constant. i.e. Build 3 CIs here, the first with $\bar{y} = 10, \sigma = 2, n = 10$ and 90% confidence level.
2. n is 10, 20, 50, 100, keeping everything else constant (use 95% confidence level).
3. $\sigma = 2, 4, 6, 8, 10$, keeping everything else constant (use 95% confidence level).

v. Twp-Sided Confidence Intervals

point estimate ± c × standard deviation of the estimator (standard error)

$$\bar{y} \pm c \times \frac{\sigma}{\sqrt{n}}$$

Where c × standard deviation of the estimator is referred to as the margin of error/sampling error/sampling allowance.
$c = Z_{\frac{1+p}{2}}$ where $Z_{\frac{1+p}{2}}$ is the $\frac{1+p}{2}$ quantile of the standard normal distribution.

vi. Summary of Symboles

	Parameter	Point Estimate	Point Estimator	Interval Estimate	Interval Estimator
	θ	$\widehat{\theta}$ (#)	$\widetilde{\theta}$ (R.V.)	$[L(y), U(y)]$ ([#],[#])	$[L(y), U(y)]$ (random interval)
$G(\mu, \sigma)$ σ is known	μ (pop mean)	$\widehat{\mu} = \bar{y}$	$\widetilde{\mu} = \bar{Y}$	$[\bar{y} - c\frac{\sigma}{\sqrt{n}}, \bar{y} + c\frac{\sigma}{\sqrt{n}}]$	$[\bar{Y} - c\frac{\sigma}{\sqrt{n}}, \bar{Y} + c\frac{\sigma}{\sqrt{n}}]$
$Bin(n, \theta)$	θ (pop prop)	$\widehat{\theta} = \frac{y}{n}$	$\widetilde{\theta} = \frac{Y}{n}$	$\frac{y}{n} \pm c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$	$\frac{Y}{n} \pm c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}$

Approximate Pivotal Quantities
$\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$

$$P\left(-c \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le c\right) = p$$

vii. Approximate Condifence Intervals for Binomial

If $Y \sim Bin(n,\theta)$, then by CLT

$$\frac{Y-n\theta}{\sqrt{n\theta(1-\theta)}} \sim G(0,1)$$

Replace θ in the denominotor by the maximun likelihood estimator $\tilde{\theta} = \frac{Y}{n}$, then divide top and botton by n, we obtain

$$Q_n = Q_n(Y;\theta) = \frac{Y-n\theta}{\sqrt{n\tilde{\theta}(1-\tilde{\theta})}} = \frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}} \sim G(0,1)$$

For approximate 95% confidence interval for θ where $\hat{\theta} = \frac{y}{n}$, we have

$$0.95 \sim P(-1.96 \le Q_n \le 1.96) = P(-1.96 \le \frac{\tilde{\theta}-\theta}{\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}} \le 1.96)$$

$$= P(\tilde{\theta} - 1.96\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}} \le \theta \le \tilde{\theta} + 1.96\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{n}}$$

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

Gaussian	Binomial
Pivotal quantity: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$	**Approx.** pivotal quantity: $\frac{\tilde{\theta}-\theta}{\sqrt{\tilde{\theta}(1-\tilde{\theta})/n}} \sim G(0,1)$ approx.
1. Quantiles: $P\left(-c \le \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \le c\right) = p$	1. Quantiles: $P\left(-c \le \frac{\tilde{\theta}-\theta}{\sqrt{\tilde{\theta}(1-\tilde{\theta})/n}} \le c\right) \approx p$
2. Rearrange: $P\left(\bar{Y} - c\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{Y} + c\frac{\sigma}{\sqrt{n}}\right) = p$	2. Rearrange: $P\left(\tilde{\theta} - 1.96\left[\frac{\tilde{\theta}(1-\tilde{\theta})}{n}\right]^{\frac{1}{2}} \le \theta \le \tilde{\theta} + 1.96\left[\frac{\tilde{\theta}(1-\tilde{\theta})}{n}\right]^{\frac{1}{2}}\right)$
3. CI: $\left[\bar{y} - c\frac{\sigma}{\sqrt{n}}, \bar{y} + c\frac{\sigma}{\sqrt{n}}\right]$ $P(-c \le Z \le c) = p,\ c = z_{\frac{1+p}{2}}$ $Z \sim G(0,1)$	3. **Approx.** CI: $\left[\hat{\theta} - c\left[\frac{\hat{\theta}(1-\hat{\theta})}{n}\right]^{\frac{1}{2}}, \hat{\theta} + c\left[\frac{\hat{\theta}(1-\hat{\theta})}{n}\right]^{1/2}\right]$ $\hat{\theta} = \frac{y}{n}$

viii. Choose Sample Size

We need to decide how large a sample to collect based on:

- how narrow we would like our confidence interval to be
- how much we can afford to spend $$

Example:

Suppose that in order to estimate θ, the proportion of units in a large population who have a specific charactieristic, we plane to select n units randomly.
Suppose that we intend to use the approximate 95$ confidence interval for θ given by

$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

It's same to

$$2(1.96)\left[\frac{\hat{\theta}(1-\hat{\theta})}{n}\right]^{1/2}$$

A criterion that is widely used is to choose the sample size, n, large enough so that the width/length of the approximate 95% confidence interval is no wider than 2(0.03) = 0.06.

To choose n such that $1.96\left[\frac{\hat{\theta}(1-\hat{\theta})}{n}\right]^{1/2} \leq 0.03$, we need to solve for n.

$$n \geq \left(\frac{1.96}{0.03}\right)^2 \hat{\theta}(1-\hat{\theta})$$

Theorem: Better Safe Than Sorry
Let's look at the margin of error, in general

$$m = E = c \times \left[\frac{\hat{\theta}(1-\hat{\theta})}{n}\right]^{1/2}$$

Since $0 < \hat{\theta} < 1$, for the most conservative assumption?
We want to maximize $\frac{\hat{\theta}(1-\hat{\theta})}{n}$, take the derivative, set it equal to 0, and solve

$$\frac{dE}{d\hat{\theta}} = \frac{(1-\hat{\theta}) + \hat{\theta}(-1)}{n} = \frac{1-2\hat{\theta}}{n}$$

Setting this equal to 0, we get $\hat{\theta} = 0.5$.
Back to the example, with E = 0.03 and $\hat{\theta} = 0.5$

$$n \geq \left(\frac{1.96}{0.03}\right)^2 \times 0.5 \times 0.5 = 1067.1 \text{ (Round up)} \nearrow$$

Therefore, if we choose $n = 1,068$, then the approximate 95% confidence interval for θ will have width less than 0.06 for all value of $\hat{\theta}$.
Example:

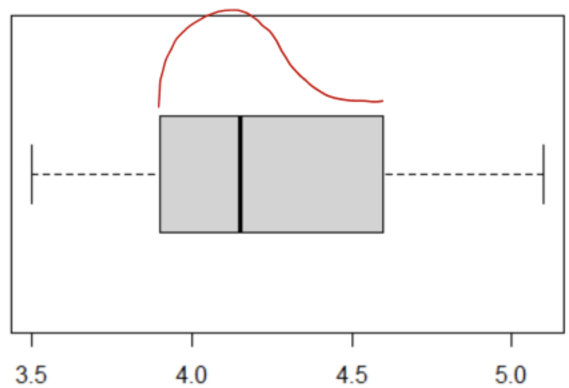
Suppose the weight of a Canada geese can be modelled with a Gaussian distribution, $G(\mu, \sigma = 0.5)$.
A sample of 10 Canada geese is taken and their weights in kgs are listed below:

$$4.4, 4.1, 3.5, 4.6, 5.1, 4.7, 3.8, 4, 3.9, 4.2$$

1. Solve for the five number summary.
2. Plot a boxplot.
3. Solve for a 90% confidence interval for the unknown average population weight, μ.
4. Solve for a 95% confidence interval for tje proportion of Canada geese that weigh more than 4.3 kgs.
5. What sample size would be large enough so that the width/length of the approximate 95%confidence interval for μ is no wider than 2(0.2)=0.4?

1. 3.5, 3.8,3.9 ,4, 4.1,4.2 ,4.4, 4.6,4.7 ,5.1

Min	3.5
q(0.25)	3.85
q(0.5)	4.15
q(0.75)	4.6
Max	5.1



2.

3. A 90% confidence interval for μ is

$$\bar{y} = 4.23, \sigma = 0.5, n = 10, 90\% = 100p\% (p = 0.9)$$

$$\begin{aligned}\bar{y} \pm c(\sigma/\sqrt{n}) &= \bar{y} \pm Z_{\frac{1+p}{2}} \times \frac{\sigma}{\sqrt{n}} \\ &= 4.23 \pm 1.645 \times \frac{0.5}{\sqrt{10}} \\ &= 4.23 \pm 0.26 \\ &= [3.97, 4.49]\end{aligned}$$

   Therefore, we are 90% confident that the average weight of a Canada goose is between 3.97 and 4.49 kgs.

4. A 95% confidence interval for θ
   $\hat{\theta} = \frac{4}{10}, n = 10, 95\% = 100p\% (p = 0.95), Z_{\frac{1+p}{2}} = 1.96$
   $\hat{\theta} = \pm z_{0.975} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} = 0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times 0.6}{10}} = 0.4 \pm 0.30364 = [0.096, 0.704]$
   Therefore, we are approxi 95% confident that the true (but unknown) proportion of geese that weigh more than 4.3 kgs is between 9.6% and 70.4%.

5. Worst case senario: $\hat{\theta} = 0.5$

$$n \geq \left(\frac{1.96}{0.2}\right)^2 \times 0.5 \times 0.5 = 24.01 \text{ (Round up)} \nearrow$$

   Therefore, if we choose $n = 25$, then the approximate 95% confidence interval for μ will have width less than 0.4 for all values of $\hat{\theta}$.

ix. Likelihood-Based Confidence Intervals

$$\Lambda = \Lambda(\theta : \widetilde{\theta}) = -2[r(\theta)] = -2 \log \left[\frac{L(\theta)}{L(\widetilde{\theta})}\right] = 2[l(\widetilde{\theta}) - l(\theta)]$$

where $\widetilde{\theta}$ is the MLE of θ.

Λ, which is a random variable, is called the likelihood ratio statistic.

Proposition: If $L(\theta)$ is based on $Y = (Y_1, Y_2, \cdots, Y_n)$, a ramdom sample of size n, and if θ is the true value of the scalar parameter, then the distribution of the likelihood ratio statistic Λ converges to a chi-squared distribution with 1 degree $\chi^2(1)$ of freedom as $n \to \infty$.

$$\Lambda \sim \chi^2(1), (n \to \infty)$$

Λ can be used as an approximate pivotal quantity to obtain confidence intervals of θ.

An approximate $100p\%$ confidence interval for θ, based on Λ, is determined by first finding the value of c such that $P(W \leq c) = p$, where $W \sim \chi^2(1)$

$$p = P(W \leq c) \approx P(\Lambda \leq c) = P[2[l(\tilde{\theta}) - l(\theta)] \leq c]$$

An approximate $100p\%$ confidence interval for θ for θ is given by

$$\{\theta : 2[l(\tilde{\theta}) - l(\theta)] \leq c\} = \{\theta : R(\theta) \geq e^{-c/2}\}$$

where $R(\theta) = \frac{L(\theta)}{L(\tilde{\theta})}$ is the likelihood ratio.

$$\begin{aligned} 2[l(\tilde{\theta}) - l(\theta)] &= -2[l(\theta) - l(\tilde{\theta})] \\ &= -2[\ln L(\theta) - \ln L(\tilde{\theta})] \\ &= -2[\ln \frac{L(\theta)}{L(\tilde{\theta})}] \\ &= -2 \ln R(\theta) \end{aligned}$$

$$\begin{aligned} -2 \ln R(\theta) \leq c \Leftrightarrow \ln R(\theta) &\geq -\frac{c}{2} \\ \Leftrightarrow R(\theta) &\geq e^{-\frac{c}{2}} \end{aligned}$$

Example:

Q: What's p choose for 95% confidence interval?
Let $Z \sim N(0,1).Z^2 \sim \chi^2(1)$. Consider the case where $c = 3.8416 = \chi^2_{0.95}(1) = Z^2_{0.975}$. $p = P(W \leq 3.8416) = 0.95$.

$$\{\theta : R(\theta) \geq e^{-3.8416/2}\} = 0.147 = \{\theta : R(\theta) \geq 0.147\}$$

Thus, a 14.7% likelihood interval is an approximate 95% confidence interval for θ.
Now we should round p and choose $p = 0.15$.

Example:

Q: What's the confidence coefficient for a 10% likelihood interval?
$R(\theta) \geq 0.1 \iff -2 \ln \frac{L(\theta)}{L(\tilde{\theta})} \leq -2 \ln 0.1, \{\theta : R(\theta) \geq e^{-c/2}\}$

$$\begin{aligned}P(W \le c) &= P(W \le -2\ln(0.1)), W \sim \chi^2(1)\\ &= P(|Z| \le \sqrt{-2\ln(0.1)}), Z \sim N(0,1)\\ &= P(-\sqrt{-2\ln(0.1)} \le Z \le \sqrt{-2\ln(0.1)})\\ &= 2P(Z \le \sqrt{-2\ln(0.1)}) - 1\\ &= 2P(Z \le 2.14) - 1 \qquad \approx 2(0.9834) - 1 = 0.9668\end{aligned}$$

Therefore, a 10% likelihood interval is an approximate 96.68% confidence interval for θ.

Note: A smaller % of the likelihood interval $\leftrightarrow$ A higher % of the confidence interval.

For data y from a $Binomial(n,p)$ distribution, two approximate 95% confidence intervals for θ:

1. Central Limit Theorem:
$$\hat{\theta} \pm 1.96\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$
where $\hat{\theta} = \frac{y}{n}$ (sample proportion)
2. a 15% likelihood interval.

For a random sample $y_1, y_2, \cdots, y_n$ from the $Poisson(\theta)$ distribution, we have two approximate 95% confidence intervals for θ:

1. $$\theta \pm 1.96\sqrt{\frac{\theta}{n}}$$
where $\hat{\theta} = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, and
2. a 15% likelihood interval.

Example:

Produce 2 CI's for each case here and compare the two intervals for:
For large samples, the confidence interals for λ can be approximated using:
$$\left(\hat{\theta} \pm Z_{\alpha/2}\sqrt{\frac{\hat{\theta}}{n}}\right)$$
where $\hat{\theta} = \frac{\lambda}{n}$ is the sample proportion and $Z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution.

1. $n = 30$ and $\hat{\theta} = 0.2$
$\lambda = \hat{\theta}n = 6$
For a 95% confidence interval, $Z_{0.025} = 1.96$.
$$\left(0.2 \pm 1.96\sqrt{\frac{0.2}{30}}\right) = (0.2 \pm 0.15) = (0.05, 0.35)$$
2. $n = 30$ and $\hat{\theta} = 0.7$
$\lambda = \hat{\theta}n = 21$
For a 95% confidence interval, $Z_{0.025} = 1.96$.
$$\left(0.7 \pm 1.96\sqrt{\frac{0.7}{30}}\right) = (0.7 \pm 0.2994) = (0.4006, 0.9994)$$

x. Chi-squared distribution

Recall:
Theorem 35:
If c is a value s.t. $p = P(W \leq c)$ where $W \sim \chi^2(1)$, then an approximate $100p\%$ confidence interval for θ is the likelihood interval $\{\theta : R(\theta) \geq e^{-c/2}\}$ Approx.CI $\rightarrow$ LI
Theorem 34:
A 100p% likelihood interval is an approximate $100p\%$ CI where $q = 2P(Z \leq \sqrt{-2\ln p}) - 1$ and $Z \sim N(0,1)$. LI $\rightarrow$ Approx.CI
Chi-squared distribution is a distribution that is frequently used in methods of estimation and hypothesis testing in the Chi-squared distribution.
The Chi-squared distribution is a special ase of a Gamma distribution, which is used extensively for modeling lifetime data(time to failure).

1. Review: The Gamma Function

Def :
The Gamma functiopn is defined as:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx, \alpha > 0$$

Properties:

- $\Gamma(\alpha) = (\alpha - 1)!$ for $\alpha = 1, 2, 3, \cdots$
- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\Gamma(0.5) = \sqrt{\pi}$

2. Chi-Squared Distribution

Suppose that X is a random variable with pdf given by:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}, x > 0, k = 1, 2, 3, \cdots$$

Then X has a chi-squared distribution with k degrees of freedom, denoted by $X \sim \chi^2(k)$.

Different values of k give different chi-squared distributions.

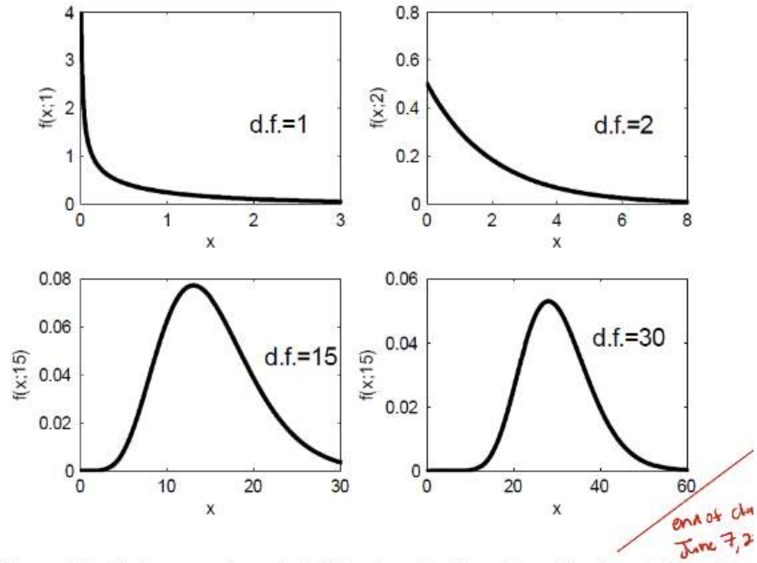


Figure 4.5: **Chi-squared probability density functions for $k = 1, 2, 15, 30$**

3. Relationship Between the Normal and Chi-squared distributions

If $Z \sim N(0, 1)$, then $W^2 \sim \chi^2(1)$.
This result aloows us to calculate probabilities associated with the random variable $W \sim \chi^2(1)$ using Normal tables:

$$P(W \leq w) = P(|Z| \leq \sqrt{w}) = 2P(Z \leq \sqrt{w}) - 1$$
$$P(W \geq w) = P(|Z| \geq \sqrt{w}) = 2P(Z \geq \sqrt{w}) = 2(1 - P(Z \leq \sqrt{w}))$$

Another Special Case
If $X \sim \chi^2(2)$, then $X \sim \mathit{Exponential}(2)$

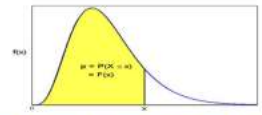
$$\begin{aligned} f(x; 2) &= \frac{1}{2^{2/2}\Gamma(2/2)} x^{2/2-1} e^{-x/2}, x > 0 \\ &= \frac{1}{2} e^{-x/2}, x > 0 \end{aligned}$$

$$P(X \leq x) = 1 - e^{-x/2} \text{ and } P(X \geq x) = e^{-x/2}$$

4. Chi-Squared Table

A chi-squared distribution is defined by its degrees of freedom.
Once the degrees of freedom have been determined, then you simply move along the row until find the quantile looking for.

Chi-Squared Quantiles



This table gives values of x for p = P(X ≤ x) = F(x)

df\p	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.000	0.000	0.001	0.004	0.016	2.706	3.842	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.146	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.054	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.391	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.430	104.210
80	51.172	53.540	57.153	60.391	64.278	96.578	101.880	106.630	112.330	116.320
90	59.196	61.754	65.647	69.126	73.291	107.570	113.150	118.140	124.120	128.300
100	67.328	70.065	74.222	77.929	82.358	118.500	124.340	129.560	135.810	140.170

5. For Larger d.f.

Note: As the d.f (degree of freedom),k, gets larger and larger the $\chi^2(k)$ distribution behaves more and more like a $Gaussian(k, \sqrt{2k})$ random variable.
(As we'll see in a bit, the chi-squared is a sum of independent random variables)

xi. Distribution of a sum of independent chi-squared random variables

Theorem :
Suppose $W_1, W_2, \cdots, W_k$ are independent random variables and $W_i \sim \chi^2(k_i)$, for $i = 1, 2, \cdots, n$.

$$S = \sum_{i=1}^{n} W_i \sim \chi^2(\sum_{i=1}^{n} k_i)$$

Note: This can be shown using the moment generating function.
Theorem :
If $Z_1, Z_2, \cdots, Z_n$ are independent and identically distribution (i.i.d) N(0,1) random variables,

$$S = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

Corollary: If $X_1, X_2, \cdots, X_n$ are independent identically distributed (i.i.d) N(μ, σ^2) random variables, then:

$$S = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

1. Summary of Chi-Squared Distribution

- If $W \sim \chi^2(k)$, then $E(W) = k$ and $Var(W) = 2k$.
- If $W_1, W_2, \cdots, W_k$ are independent random variables with $W_i \sim \chi^2(k_i)$, then $S = \sum_{i=1}^{n} W_i \sim \chi^2(\sum k_i)$.
- If $W = Z^2 \sim \chi^2(1)$, then $P(W \leq w) = P(|Z| \leq \sqrt{w}) = 2P(Z \leq \sqrt{w}) - 1$, where $Z \sim G(0,1)$.
- If $Z_1, Z_2, \cdots, Z_n \sim G(0,1)$, then $S = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$.
- If $X \sim \chi^2(2)$, then $X \sim Exponential(2)$ and $P(X \leq x) = 1 - e^{-x/2}$.

xii. Student's t-distribution

The Student t distribution (t distribution for short) has probability density function (pdf)

$$f(x; k) = c_k \cdot \left(1 + \frac{x^2}{k} \right)^{-\frac{k+1}{2}}, -\infty < x < \infty \text{ and } k = 1, 2, 3, \cdots$$

where parameter k is called the "degrees of freedom" and the constant c_k is given by:

$$c_k = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)}$$

We write $T \sim t(k)$ to denote that the random variable T has a Student t distribution with k degrees of freedom.
Suppose $Z \sim G(0,1)$ and $U \sim \chi^2(k)$ are independently. Then:

$$T = \frac{Z}{\sqrt{\frac{U}{k}}} \sim t(k)$$

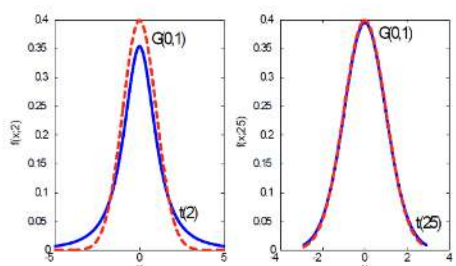


Figure 4.6: **Probability density functions for $t(k)$ (solid blue) and $G(0,1)$ (dashed red)**

1. Properties of the t-distribution

- The student t probability density function is similar to that of the G(0,1) distribution since it is unimodal and symmetric about the origin.
- For small k, the student t density has larger (thicker) tails or more area in the extreme left and right tails than the G(0,1) distribution.

- For larger k, the graph of the probability density function f(x;k) is indistinguishable from that of the G(0,1) pdf.
- In R, the cumulative distribution function (CDF), $F(x;k) = P(T \leq x;k)$ where $T \sim t(k)$, is obtained using the $pt(x,k)$ function.

Example:

Let $T \sim t(5)$,
$P(T \leq t) = 0.975$ gives $t = t_{5,0.975} = 2.571$
Compare with $Z \sim G(0,1)$, $P(Z \leq z) = 0.975$ gives $z = z_{0.975} = 1.96$
Compare with $T \sim t(100)$, $P(T \leq t) = 0.975$ gives $t = t_{100,0.975} = 1.984$ (this result is close to the $Z \sim G(0,1)$ result)

2. T-distribution

This distribution is a lot like a Normal distribution.

- Unimodal
- Bell shaped
- Symmetric around a mean of 0

The difference is taht this distribution has thicker tails than the Normal distribution.

Theorem :

If $Z \sim G(0,1)$ and $U \sim \chi^2(n)$, then $Z\sqrt{\frac{U}{n}} \sim t(n)$.

xiii. Confidence Intervals for Parameters in the $G(\mu, \sigma)$ Model

Suppose $Y \sim G(\mu, \sigma)$. We showed that if we knew σ, then for a sample with mean $\bar{y}$ a $100p\%$ CI for μ is

$$\left[\bar{y} - c\frac{\sigma}{\sqrt{n}}, \bar{y} + c\frac{\sigma}{\sqrt{n}}\right]$$

Now suppose we don't know σ. μ and σ are both unknown. Still build a 95% CI for the population mean.

We can still use the maximum likelihood estimator as a point estimator for μ.

There are two options for the CI:

1. Maximum likelihood estimate

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

with estimator $E(\tilde{\sigma}^2) \neq \sigma^2$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

2. Sample variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

with estimator

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

We prefier S^2 because $E(S^2) = \sigma^2$. (Unbiased)

Since $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$, which uses $\bar{Y}$ as an estimator for μ, what happens if we replace σ with its estimator S?

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim ?$$

Important result: If $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, then $U = \frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$.(Proof in later notes)

Example:

Driving Speeds Example

The following data gives the speed (in km/hr), as measured by radar of 10 cars traveling on a major highway:

$$112, 116, 118, 121, 112, 108, 104, 115, 121, 106$$

It seems reasonable to assume that the speeds are normally distributed with model $G(\mu,\sigma)$ for $i = 1, 2, \ldots, 10$. independent.

$$\sum y_i = 1133, \sum y_i^2 = 128691, \bar{y} = 113.3$$

$$S^2 = \frac{1}{9}\left(128691 - \frac{1133^2}{10}\right) = 35.7889$$

We wish to estimate the parameter, μ, the mean driving speed on this major highway.
A point estimate for μ is $\bar{y} = 113.3$.
An interval estimate is given by a 95% CI. To obtain the confidence interval, we use the pivotal quantity

$$T = \frac{\bar{Y}-\mu}{\frac{S}{\sqrt{10}}} \sim t(9)$$

$$\bar{y} \pm t_{(1+p)/2,n-1}\frac{S}{\sqrt{n}}$$

From the t-tables, we see $P(T \leq 2.2622) = (1 + 0.95)/2 = 0.975$.
Therefore, a 95% CI for μ is

$$\bar{y} \pm 2.2622\frac{S}{\sqrt{10}} = 113.3 \pm 2.2622\frac{\sqrt{35.7889}}{\sqrt{10}} = 113.3 \pm 4.2796 = [109.0204, 117.5796]$$

xiv. Sample size determination for a given confidence interval width

If σ was known, we could esaily determine the sample size necessary to achieve a certain level of precision. For a 95% confidence interval, with σ known, the margin of error was given by

$$E = m = z_{(1+p)/2}\frac{\sigma}{\sqrt{n}} = z_{0.975}\frac{\sigma}{\sqrt{n}} = 1.96\frac{\sigma}{\sqrt{n}}$$

.

$$n = (1.96\frac{\sigma}{m})^2 \text{ (n should be rounded up to the nearest integer)}$$

For the case σ is unknown, we can use $t_{(1+p)/2,n-1}$ instead of $z_{(1+p)/2}$.

$$n = (t_{(1+p)/2,n-1}\frac{S}{m})^2$$

(BUT WE DON'T KNOW n)

To be safe, we would normally choose n a little larger than this formula suggests,

1. we don't know the actual value of σ
2. The quantile of the t-distribution is larger in value than the quantile of the Gausian distribution.

xv. Confidence Intervals for the Gaussian Variance with Mean Unknown

Suppose that we have a random sample $Y_1, Y_2, \ldots, Y_n$ from a $G(\mu, \sigma)$ model. We use the estimator S^2 to build a confidence interval for the parameter σ^2.

Theorem :

Suppose $Y_1, Y_2, \ldots, Y_n$ are independent random varaibles with common $G(\mu, \sigma)$ distribution and suppose S^2 is the sample variance. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2(n-1)$$

1. Degrees of Freedom

$df = n - 1$ is the degrees of freedom of the χ^2 distribution.

Note: Although it can be shown that $(Y_i - \bar{Y})$ is Normally distributed, the random variables $(Y_i - \bar{Y}), i = 1, 2, \ldots, n$ are NOT independent.

We know that

$$\sum_{i=1}^{n}(Y_i - \bar{Y}) = 0 \Rightarrow \sum_{i=1}^{n-1}(Y_i - \bar{Y}) + (Y_n - \bar{Y})$$

.

there are only $n-1$ independent terms that are linearly independent, "free" to vary.

$$Y_n - \bar{Y} = -\sum_{i=1}^{n-1}(Y_i - \bar{Y})$$

2. Confidence Interval for the Variance

Recall:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2(n-1)$$

We can use

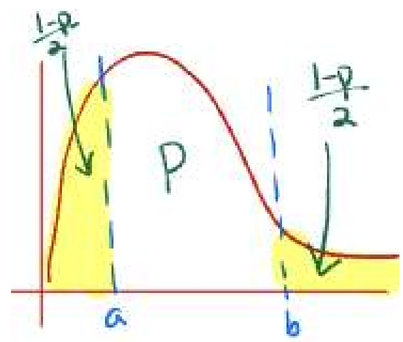
$$U = \frac{(n-1)S^2}{\sigma^2}$$

as a pivotal quantity in order to construct a 100p% confidence interval for the parameter σ^2.

For a two-sided, equal-tailed confidence interval, we need to find a and b using the chi-squared tables, such that

$$P(U \le a) = P(U \ge b) = \frac{1-p}{2}$$

where $U \sim \chi^2(n-1)$.



Since

$$\begin{aligned} P &= P(a \le U \le b) \\ &= P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) \\ &= P\left(\frac{(n-1)s^2}{b} \le \sigma^2 \le \frac{(n-1)s^2}{a}\right) \end{aligned}$$

Thus. a $100p\%$ confidence interval for σ^2 is

$$\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right] = [L(y), U(y)]$$

where s^2 is the observed sample variance. It's NOT symmetric about s^2. The point estimate of σ^2.

The 100% confidence interval for σ is

$$\left[\sqrt{\frac{(n-1)s^2}{b}}, \sqrt{\frac{(n-1)s^2}{a}}\right]$$

Example:

The professor weighs their sample of 30 donuts and finds the sample variance is $s^2 == 0.311$.
Let Y denote the weight of a randomly selected Monster Donut and assume $Y \sim G(\mu, \sigma)$ where μ and σ are unknown.
Construct a 95% confidence interval for σ^2 based on these data.

$$P(U \leq a) = \frac{1-p}{2} = 0.025, U \sim \chi^2(29)$$
$$P(U \geq b) = 0.025$$
$$P(U \leq b) = 0.975$$

Frp, the chi-squared table with 29 degrees of freedom, we find that $a = 16.047$ and $b = 45.722$.
Thus, the resulting 95% confidence interval for σ^2 is given by:

$$\left(\frac{(n-1o)s^2}{b}, \frac{(n-1)s^2}{a}\right) = \left(\frac{29 \times 0.311}{45.722}, \frac{29 \times 0.311}{16.047}\right) = (0.197, 0.562)$$

CI for μ (σ known)	**CI for μ (σ unknown)**	**CI for σ^2**
Pivotal quantity: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$	Pivotal quantity: $\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$	Pivotal quantity: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
1. Quantiles: $P\left(-c \leq \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq c\right) = p$	1. Quantiles: $P\left(-c \leq \frac{\bar{Y}-\mu}{\frac{S}{\sqrt{n}}} \leq c\right) = p$	1. Quantiles: $P\left(a \leq \frac{(n-1)S^2}{\sigma^2} \leq b\right) = p$
2. Rearrange: $P\left(\bar{Y} - c\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{Y} + c\frac{\sigma}{\sqrt{n}}\right) = p$	2. Rearrange: $P\left(\bar{Y} - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{Y} + c\frac{S}{\sqrt{n}}\right) = p$	2. Rearrange: $P\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right) = p$
3. CI: $\left[\bar{y} - c\frac{\sigma}{\sqrt{n}}, \bar{y} + c\frac{\sigma}{\sqrt{n}}\right]$ $P(-c \leq Z \leq c) = p$, $c = z_{\frac{1+p}{2}}$ $Z \sim G(0,1)$	3. CI: $\left[\bar{y} - c\frac{s}{\sqrt{n}}, \bar{y} + c\frac{s}{\sqrt{n}}\right]$ $P(-c \leq T \leq c) = p$, $c = t_{\frac{1+p}{2};n-1}$ $T \sim t(n-1)$	3. CI: $\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$ $P(U \leq a) = P(U > b) = \frac{1-p}{2}$ $U \sim \chi^2(n-1)$

IV. Tests of Hypotheses

Statistical tests of hypotheses are conducted in a similar manner to a North American criminal trial.
We begin our process by specifying a single "default" hypothesis.
This hypothesis is called the **null hypothesis** and is denoted by H_0.
There is also an alternative hypothesis denoted by H_1.
As the name implies, this is the alternative to the null hypothesis.

Recall: We're trying to see if the professor's multiple-choice questions are truly randomized.

- Let Y equal the number of questions with the answer 'A' out of a randomly chosen 25 questions.
- We assume Y has a Binomial distribution with $n = 25$, and let θ be the probability of getting 'A'.
- If the questions are rabdomized, then $\theta = 0.5$, otherwise $\theta \neq 0.5$.

- Null hypothesis: $H_0 : \theta = 0.5$
- Observed values consistent with H_0: values close to 12.5
- We need a way to quantify the strength of evidence aagainst H_0!

i. Test Statistics and Discrepancy Measures

This motivates a very important definition

Def :
A test statistic or discrepancy measure is a function of the data $D = g(Y)$ that is constructed to measure the degree of "agreement" between the data Y and the null hypothesis H_0.

Important: D is a function of Y and is thus a random variable. Once we observe $Y = y$, then the observed value of D is dennoted $d = g(y)$.

We usually define D so that $d = 0$ represents the best possible agreement between the data and H_0.

Larger values of d indicate poorer agreement between the data and H_0.

$$D(Y) = |Y - E(Y)| = |Y - 12.5|$$

Smaller observed values of D provide less evidence againist H_0.

Larger observed values of D provides more evidence against H_0.

We have evidence against the null hypothesis $H_0 : \theta = 0.5$

Key concept: we suppose H_0 is true and try to quantify the strength of evidencce provided by our observed data.

Example:

Suppose in our 25-question Quiz we observe 10 'A' answers, and so the observed value of $D = |Y - 12.5|$ is

$$d = |10 - 12.5| = 2.5$$

Let's compute $P(D = 2.5)$ assuming $\theta = 0.5$. We can start with

$$P(Y = 10) = \binom{25}{10}\theta^{10}(1-\theta)^{15} = \binom{25}{10}0.5^{25} = 0.097$$

We would be less suprised if $d = |Y - 12.5| < 2.5$.

$$d = |10 - 12.5| = 2.5 \Rightarrow Y = 10, 11, 12, 13, 14$$

$$P(D \geq 2.5; H_0) = P(|Y - 12.5| \geq 2.5)$$

where $Y \sim Bin(25, 0.5)$. Therefore,

$$\begin{aligned} P(Y \leq 10) + P(Y \geq 15) &= 1 - P(11 \leq Y \leq 14) \\ &= 1 - \sum_{y=11}^{14} \binom{25}{y} 0.5^{25} \\ &= 0.4244 \end{aligned}$$

In other words, the probability of ovserving a discrepancy measure D at least as large as that which we actually observed(2.5) is 0.424, or 42.4%. **Important Note:** Instead just the probability we can not prove that the professor's multiple-choice questions are not truly randomized.

The probability that we ahve calculated is called the **p-value** of the observed data.

ii. The p-value

Def :
The p-value of the test of hypothesis H_0 using test statistic D is $P(D \geq d; H_0)$.

- The p-value represents the porbability of observing a value of the test statistic greater than or euqal to the observed value of the test statistic assuming H_0 is true.
- It is a measure of the level of evidence against H_0 based on the observed data.
- So, the smaller the p-value, the more evidence we have against H_0, or the less the data supports the claim that the null hypothesis is true.

Important: we want $P(D \geq d; H_0)$, and not $P(D = d; H_0)$. We want to know the probability that, if the null hypothesis were true, we'd see something at least as extreme/unusual as what was actually observed.

1. P-Value Interpretation

Note: p-values are super important but are also one of the most commonly misunderstood concepts in statistics!
Key concept: p-values are used for communicating the strength of evidence against H_0, and therefore we ask what is the probability of observomg data at least as suprising as those actually obserbed?

2. Steps in a Statistical Hypothesis Test

1. Specify the null hypothesis H_0, to be testwed using some data, Y.
2. Adopt a test statistic/discrepancy measure, $D = D(Y)$, for which large values of D are less consistent with H_0. Let $d = D(y)$ be the corresponding observed value of D.
3. Calculate the observed p-value where

$$\text{p-value} = P(D \geq d; H_0)$$

4. Draw a conclusion based on the p-value.

Example:

Suppose our 25-question quiz has 21 A' answers. The p-value for a test of $H_0 : \theta = P(A) = 0.5$ is:

$$P(D \geq |21 - 12.5|; H_0) = P(D \geq 8.5; H_0) = 0.0009$$

3. Guidelines for interpreting the P-Value

P-Value	**Interpretation**
$P \leq 0.001$	Very strong evidence against H_0
$0.001 < P \leq 0.01$	Strong evidence against H_0
$0.01 < P \leq 0.05$	Evidence against H_0
$0.05 < P \leq 0.10$	Weak/Some evidence against H_0
$P > 0.10$	No evidence against H_0

iii. Errors

In some instance, we may be required to make a decision - reject H_0 or fail to reject H_0. Based on a set level of significance(α). It's possible our decision could be wrong.

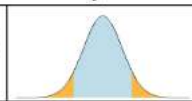
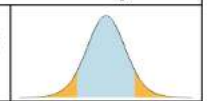
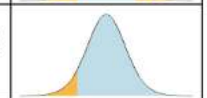
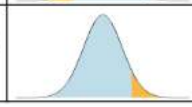
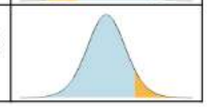
$$\alpha = P(\text{Type I Error}) = P(\text{Reject } H_0 | H_0 \text{ is true})$$

$$\beta = P(\text{Type II Error}) = P(\text{Fail to Reject } H_0 | H_0 \text{ is false})$$

iv. Hypothesis Tests for Gaussian Parameters

Example:

A 'large' tea at Tim Hortons should contain 590mL of liquid. Suppose the company wanted to check that a store was serving the correct amount and ordered 25 cups of tea.
We might suppose $Y \sim G(\mu, \sigma^2)$ denotes the volume of tea in a randomly chosen cup. Because we sould be getting, on average, 590ml of tea, we might test the null hypothesis $H_0 : \mu = 590$.

H_A	P-Value			
	Testing $H_0: \mu = \mu_0$ (σ known)		Testing $H_0: \mu = \mu_0$ (σ unknown)	
$\mu \neq \mu_0$	$P(D \geq d; H_0 \text{ is true})$ $= P(\lvert Z \rvert \geq d)$		$P(D \geq d; H_0 \text{ is true})$ $= P(\lvert T \rvert \geq d)$	
$\mu < \mu_0$	$P(D \leq d; H_0 \text{ is true})$ $= P(Z \leq d)$		$P(D \leq d; H_0 \text{ is true})$ $= P(T \leq d)$	
$\mu > \mu_0$	$P(D \geq d; H_0 \text{ is true})$ $= P(Z \geq d)$		$P(D \geq d; H_0 \text{ is true})$ $= P(T \geq d)$	

1. P-Value and Confidence Intervals

Even if our p-value is small, suggestion we have evidence against H_0, the size of the p-value does not imply hwo 'wrong' H_0 is. It's possible that H_0 is only slightly wrong.

The p-value just tells us how suprised we'd be by these date if the null hypothesis were true.

A confidence interval however does indicate the magnitude and direction of the departure from H_0.

If strong evidence against H_0 is found in a particular direction then this might suggest conducting further experiments to investigate this evidence.

Example:

Mr. Chips claims that their potato chip company produces bags of chips containing at least 200g per bag on average. A quanlity control engineer believes the machine responsible for filling the bags is actually under filling them and whies to test their hypothesis. They take a random sample of 20 bags of chips from the production line and find the average weight to be 198.3g with a variance of 9.27. The weights are auumed to be normally distributed.

$n = 20, \bar{y} = 198.3, s^2 = 9.27, \mu_0 = 200, \sigma^2 = 9.27$

Q: Is there sufficient evidence to conclude that the bags are being under filled, on average

1. Hypothesis: In this case, we want to test
$$H_0 : \mu = 200 \text{ vs } H_1 : \mu < 200$$
Note: This is a ONE-SIDED test.
2. Observed test statistic: In this case, we have a Normal parent population with the population variance unknown.
$$D = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{198.3 - 200}{\sqrt{9.27}/\sqrt{20}} = -2.497$$
3. p-value:
$$P(D \leq d; H_0) = P(T \leq -2.497) \text{ where } T \sim t_{n-1}$$
$$P(T \leq -2.497) = pt(-2.497, df = 19) = 0.01093916$$
4. Conclusion: There is evidence against $H_0, \mu = 200$ based on the observed data (evidence of under-filling)

2. Relationship Between Tests of Hypothesis and Confidence Intervals

Suppose $y_1, y_2, \ldots, y_n$ is an observed random sample from the $G(\mu, \sigma^2)$ distribution. Suppose we test $H_0 : \mu = \mu_0$ v.s. $H_A : \mu \neq \mu_0$ (two-sided test).

We will do so using a two-sided 95% confidecne interval. Simply regard the set of values contained in the 95% confidence interval as a set of plausible hypotheses. So, if

$$\mu_0 \in \left[\bar{y} - a\frac{s}{\sqrt{n}}, \bar{y} + a\frac{s}{\sqrt{n}}\right]$$

where a is the appropriate value from the t-distribution $t_{0.975}$ value with $d.f. = n-1$, this means that the p-value ≥ 0.05.

In other words:

- If our p-value is greater or equal to 0.05, then μ_0 is inside a 95% confidence interval.
- If μ_0 is inside a 95% confidence interval, then the p-value is greater or equal to 0.05.

Warning: This result only approximately holds if we use different pivotal quantities, such as using a Gaussian approximation to calculate the p-value and a chi-suared approximation to calculate the confidence interval!

Example:

Two weight scales:

- Scale 1: $p = 0.423$ and a 95% CI was $[0.982, 1.03]$
- Scale 2" $p = 0.0064$ and a 95% CI was $[0.969, 0.993]$

If we knew the 95% CI for the first scale was $[0.982, 1.03]$, we would know the p-value for a test of $H_0 : \mu = 1$ would be greater than 0.05.

If we knew the 95% CI for the second scale was $[0.969, 0.993]$, we would know the p-value for a test of $H_0 : \mu = 1$ would be less than 0.05.

3. Test for σ^2 in the Gaussian Model

In general, a 100p% CI for σ^2 based on a sample of size n with sample variance s^2 is given by:

$$\left[\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right]$$

which made use of the pivotal quantity

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

4. Summary of Testing $H_0 : \mu = \mu_0$ (σ unknown)

1. Test $H_0 : \mu = \mu_0$ based on the observed random sample $y_1, y_2, \ldots, y_n$ from the $G(\mu, \sigma^2)$ distribution.
2. Use the test statistic $D = \frac{|\bar{Y} - \mu_0|}{S/\sqrt{n}}$ with observed value $d = \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}$.
3. Calculate p-value $= 2[1 - P(T \leq d)]$ where $T \sim t_{n-1}$.
4. Draw conclusions based on the p-value

Bonus fact: if p-value $< 1 - q$, then a 100q% confidence interval for μ will not contain μ_0, and vice versa.

5. Summary of Testing $H_0 : \sigma^2 = \sigma_0^2 (\mu$ unknown)

To test $H_0 : \sigma^2 = \sigma_0^2$ in the $G(\mu, \sigma)$ model:

1. Draw a sample of size n with sample vairance s^2.
2. Use the test statistic $U = \frac{(n-1)s^2}{\sigma_0^2}$ with observed value $u = \frac{(n-1)s^2}{\sigma_0^2}$.
3. Compute $P(U \le u)$ for $U \sim \chi^2_{n-1}$.
4. (a) If $P(U \le u) < 0.5$, then p-value $= 2[P(U \le u)]$.
   (b) If $P(U \ge u) < 0.5$, then p-value $= 2[P(U \ge u)]$.
5. Draw conclusions based on the p-value.

v. Likelihood Ratio Tests of Hypotheses - One parameter

We derive a more general method for conducting hypothesis tests for any distribution.

1. Likelihood Ratio Test

Recall: We introduced the **likelihood ratio statistic:**

$$\Lambda = -2 \log \left[\frac{L(\theta)}{L(\tilde{\theta})} \right] = -2 \log \left[\frac{L(\theta; Y)}{L(\hat{\theta}; Y)} \right]$$

where $\tilde{\theta} = \tilde{\theta}(Y)$ is the MLE of θ.
For large n, $\Lambda \sim \chi^2_1$ approximatelym and we used this result to demonstrate a link between likelihood and confidence intervals.

We can also use this result to construct an approximate hypothesis test of $H_0 : \theta = \theta_0$ v.s. $H_A : \theta \ne \theta_0$.

Principles of constructing a discrepancy measure:

1. If H_0 is true, we know the (approximate) distribution of the discrepancy measure.
2. If the data are inconsistent with H_0, the observed value of the discrepancy measure should be large.

V. Gaussian response models

i. Bivariate Data

So far, the vast majority of what we've learned has been in the context of univariate data.

We are often interested in studying bivariate data, where two variates are measured per unit.

In particular, we can study the **relationship** between the two variates.

Review: Sample Correlation

The linear relationship between two variates:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}, \quad -1 \le r \le 1$$

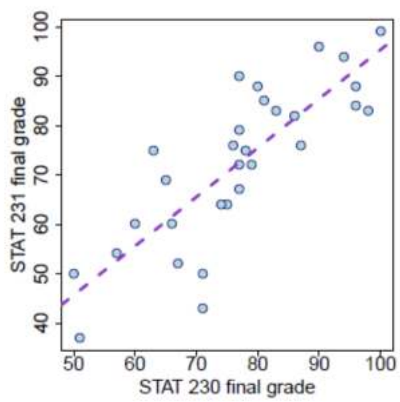
Fitting lines

A "better" line in one that minimizes the distance between itself and the data poitns, but there are multiple ways to do this.

ii. Explanatory and Response Variates

Example:

> A reasonable goal here might be to understand the relationship between STAT 230 and STAT 231 grades in terms of how much of the variation in STAT 231 grade is explained by variation in STAT 230 grade.
> Thus the STAT 231 final grade is the response variate and the STAT 230 fianal grade is the explanatory variate.



iii. Gaussian Response Models

Def :

A gaussian response model is a mode lin which a random variable, Y_i (the response variate) is assumeed to have a Gaussian distribution whose parameters depend on a known vector $x = (x_1, x_2, \dots, x_n)$ independently of explanatory variates (or covariates).

General form:

$$Y_i = G(\mu(x_i), \sigma), \quad i = 1, 2, \dots, n \text{ in dependently}$$

where the x_i's are assumed to be known. The mean of Y_i depends on the covariate x_i, but the standard deviation, σ, does not.
in another way, we can write the model as:

$$Y_i = \mu(x_i) + R_i, \quad R_i \sim G(0, \sigma)$$

iv. Simple Linear Regression Model

$$Y_i = G(\alpha + \beta x_i, \sigma), \quad i = 1, 2, \dots, n; \text{ independently}$$

where the x_i's are assumed to be known constanrts and there are three unknow parameters α, β, σ.
Note that we have assumed that the standard deviation σ does not depent on x_i.

1. The parameter β is the **slope coefficient** and it represnts the change in the mean, $\mu(x) = \alpha + \beta x$, for a unit change in x. (i.e. $\Delta x = 1$)

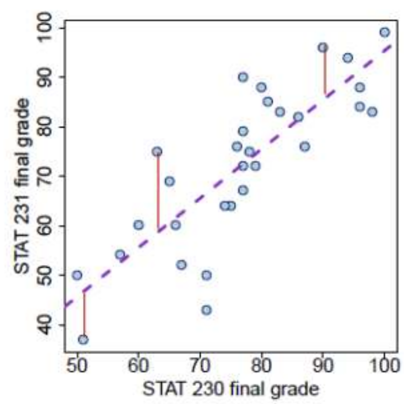
$$\begin{aligned}\mu(x+1) - \mu(x) &= \alpha + \beta(x+1) - (\alpha + \beta x) \\ &= \beta\end{aligned}$$

2. the parameter α is the **vertical intercept** and it represnts the mean response when x is 0.

$$\mu(x = 0) = \alpha + \beta \cdot 0 = \alpha$$

1. Residuals

Consider the vertical distances between a fitted line and the data. These are called residuals.



v. Least Squares Estimation

For a line $y = \alpha + \beta x$ we can define the residual for each pair of point (x_i, y_i) as $r_i = y_i - (\alpha + \beta x_i)$. To find the least squares estimates $\hat{\alpha}$ and $\hat{\beta}$, we minimize

$$g(\alpha, \beta) = \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2$$

by differentiating

$$\begin{aligned}\frac{\partial g}{\partial \alpha} &= \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-1) = 0 \\ \frac{\partial g}{\partial \beta} &= \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2 = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) = 0\end{aligned}$$

finding $\hat{\alpha}$ and $\hat{\beta}$ by solving the above two equations.

$$\begin{aligned}-2\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta} x_i) &= 0 \\ \Leftrightarrow \sum y_i - n\hat{\alpha} - \hat{\beta} \sum x_i &= 0 \\ \Leftrightarrow n\bar{y} - n\hat{\alpha} - \hat{\beta} n\bar{x} &= 0 \\ \Leftrightarrow \bar{y} - \hat{\alpha} - \hat{\beta}\bar{x} &= 0 \quad (1)\end{aligned}$$

and

$$
\begin{aligned}
-2\sum_{i=1}^{n}(y_i - \hat{\alpha} - \hat{\beta}x_i)x_i &= 0 \\
\Leftrightarrow \sum(y_i - \hat{\alpha} - \hat{\beta}x_i)x_i &= 0 \quad (2)
\end{aligned}
$$

From (1) we have $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, and substituting this into (2) we get

$$
\begin{aligned}
\sum(y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i)x_i &= 0 \\
\Leftrightarrow \sum(y_i - \bar{y})x_i - \hat{\beta}\sum(x_i - \bar{x})x_i &= 0 \\
\Leftrightarrow \hat{\beta} &= \frac{\sum(y_i - \bar{y})x_i}{\sum(x_i - \bar{x})x_i}
\end{aligned}
$$

Recall:

- $\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i(y_i - \bar{y})$
- $\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})x_i$

So, the least squares estimates are

$$
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}
$$

$$
\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}
$$

Example:

STAT 230 and STAT 231 Final grades

$$
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}}
$$

Notes:

$$
\bar{x} = \frac{2302}{30} = 76.7333, \quad \bar{y} = \frac{2167}{30} = 72.2333
$$

$$
S_{xx} = 5135.8667, \quad S_{yy} = 7585.3667, \quad S_{xy} = 5106.8667
$$

$$
\hat{\beta} = \frac{5106.8667}{5135.8667} = 0.9944
$$

$$
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 72.2333 - 0.9944 \cdot 76.7333 = -4.0667
$$

Thus, the resulting fitted line is $y = -4.0667 + 0.9944x$