VE401 Recitation 5

Fisher's Hypothesis Test

Hypotheses and Testing

Example: I am a soft drink manufacturer. You are a seller and want to buy my products. I claim that my product has $\mu \geq 500$ ml. You don't trust me and sampled 20 bottles from my batch.

In[82]:= **Data = RandomVariate[NormalDistribution[499, 2], 20]**

Out[82]= {499.728, 497.006, 498.871, 500.931, 500.157, 501.139,
  495.516, 501.89, 496.894, 497.132, 503.624, 500.357, 498.011,
  498.447, 496.074, 502.177, 498.081, 499.719, 497.657, 498.895}

In[83]:= **Mean[Data]**

Out[83]= 499.115

I claim that the samples are too **"extreme"** and you are picking bottles with small volumes by chance. Actually, more and more have larger volume. Would you believe me?

Steps:

- Figure out the ***parameter θ*** and the ***null value θ_0*** you are testing. ($\theta = \mu$, $\theta_0 = 500$ in my example)

- Set up a ***null hypothesis H_0***. It can be one of the following: ($\mu \geq 500$ in my example)

    - $\theta = \theta_0$.

    - $\theta \leq \theta_0$.

    - $\theta \geq \theta_0$.

- Figure out the distribution of the estimator of θ and thus the test statistics. ($\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ follows standard normal distribution. We take $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ as test statistics)

- Analyze data, calculate P-value to see if there is enough evidence of reject H_0.

Significance or P-value for One-Tailed Test

Intuition: How likely will we observe the data if what I claim is true? How "extreme" is the data?
$P = P[\text{getting this or even more "extreme" observation} \mid H_0 \text{ is true}]$.

Out[216]=

Hypothetical distribution of $\overline{X}$ with μ_0=500.02
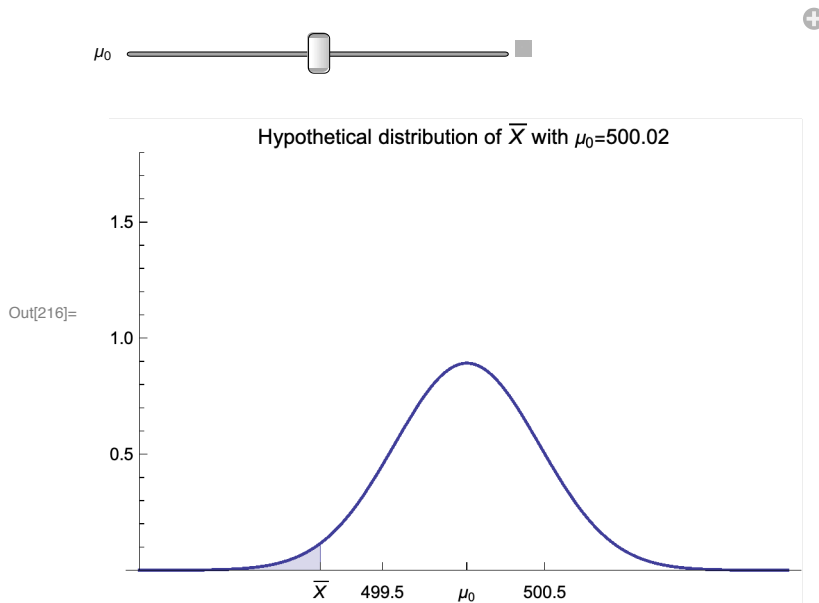
In this graph, the distribution is the hypothetical distribution of the estimator according to H_0.

Mathematical representation: Denoting the sample statistics of $\hat{\theta}$ as k, we have

$$P \text{ value} = \begin{cases} P[\hat{\theta} \leq k \,|\, \theta = \theta_0] & \text{if null hypothesis is } H_0 : \theta \geq \theta_0 \\ P[\hat{\theta} \geq k \,|\, \theta = \theta_0] & \text{if null hypothesis is } H_0 : \theta \leq \theta_0 \end{cases}$$

Example:

What is the P-value for my claim that $\mu \geq 500$ ml? Assume that we know true variance $\sigma^2 = 4$ ml^2 and sample statistics $\overline{X} = 499.115$ ml.

For $H_0 : \mu \geq 500$, we calculate (suppose we already know that $\sigma = 2$).
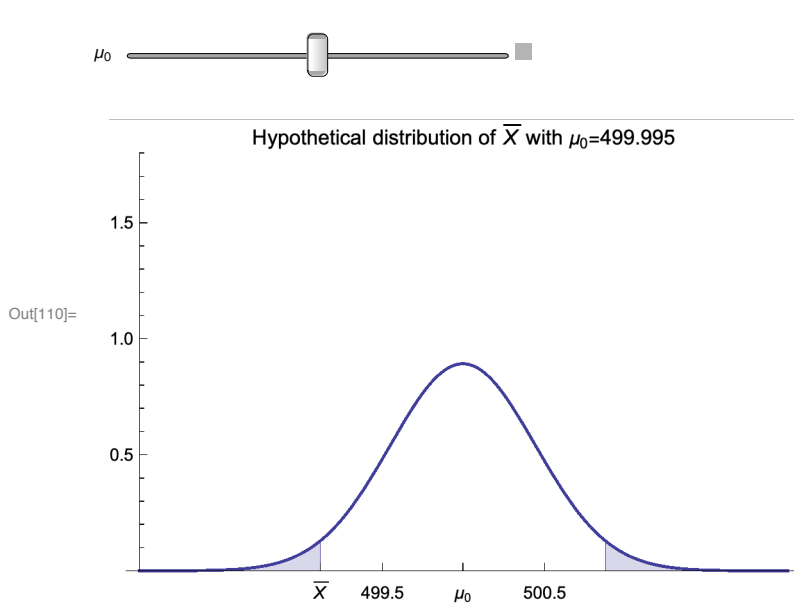
$$P[\text{getting more "extreme" observation} \,|\, H_0 \text{ is true}] = P\left[\overline{X} \leq 499.115 \,\middle|\, \mu \geq 500, \ \sigma = 2\right]$$

$$<= P\left[\overline{X} \leq 499.115 \,\middle|\, \mu = 500, \ \sigma = 2\right]$$

$$= P\left[\frac{\overline{X}-500}{2/\sqrt{20}} \leq \frac{499.115-500}{2/\sqrt{20}}\right]$$

$$= P[Z \leq -1.978]$$

$$= 0.024$$

In[111]:= `CDF[NormalDistribution[],` $\dfrac{\texttt{Mean[Data]} - 500}{2/\sqrt{20}}$`]`

Out[111]= `0.0239544`

So, there is at most 2.4% chance to obtain such "extreme" data if my claim is true. Since it is small, so we reject H_0 at the 2.4% level of significance.

Significance or *P*-value for Two-Tailed Test

Out[110]=

Hypothetical distribution of $\overline{X}$ with μ_0=499.995

Mathematical representation: Denoting the sample statistics of $\hat{\theta}$ as k, we have

$$P \text{ value } = P\big[\,\big|\hat{\theta} - \theta_0\big| \geq |k - \theta_0|\,\big|\,\theta = \theta_0\big] \quad \text{if null hypothesis is } H_0 : \theta = \theta_0$$

Example:

What is the *P*-value if my claim is $\mu = 500$ ml? Assume that we know true variance $\sigma^2 = 4$ ml^2 and sample statistics $\overline{X} = 499.115$ ml.

For $H_0 : \mu \geq 500$, we calculate (suppose we already know that $\sigma = 2$).

$$\text{getting more "extreme" observation} \,|\, H_0 \text{ is true}] = P\big[\,\big|\overline{X} - 500\big| \geq |\,499.115 - 500\,|\,\big|\,\mu = 500,\ \sigma =$$
$$= 2\,P\big[\overline{X} \leq 499.115\,\big|\,\mu = 500,\ \sigma = 1\big]$$
$$= 0.048$$

So, there is at most 4.8% chance to obtain such "extreme" data if my claim is true. Since it is small, so we reject H_0 at the 4.8% level of significance.

Neyman-Pearson Decision Theory

Hypotheses and Testing

Steps:

- Figure out the parameter θ and the null value θ_0 you are testing.

- We set up two competing hypotheses H_0 and H_1. (e.g. $H_0 : \mu \geq 500$, $H_1 : \mu \leq 499$). They can be one of the following: for some $\delta > 0$,

  - $H_0 : \theta = \theta_0$, $H_1 : |\theta - \theta_0| \geq \delta$.

  - $H_0 : \theta \leq \theta_0$, $H_1 : \theta \geq \theta_0 + \delta$.

  - $H_0 : \theta \geq \theta_0$, $H_1 : \theta \leq \theta_0 - \delta$.

- Decide α and β. Calculate the critical region and the sample size.

- Analyze data. If the sample statistics lies in the interval of H_1, we accept H_1. Otherwise, we see the whether the sample statistics lies in the critical region. If yes, then reject H_0.

Type I and Type II Errors

Intuition: Type I error is the case when my claim is falsely rejected. Type II error is the case when my claim is falsely accepted.

		In fact	
		H_0 true	H_0 false
Our choice	Accept H_0	**Good**	**Type II error**
	Reject H_0	**Type I error**	**Good**

α and Critical Region

Definition: $\alpha = P[\text{Type } I \text{ error}]$. The ***critical region*** is the interval such that if H_0 is true, the probability of statistics falling into the interval (letting us to reject H_0) is at most α. In other words, the complement of critical region is a $100\,(1 - \alpha)\,\%$ ***confidence interval*** for the parameter.

Usage: If P value $\leq \alpha$, we reject H_0. If sample statistics lies in the critical region, we reject H_0.

Example:

We set $\alpha = 0.05$. Calculate the critical region for $H_0 : \mu \geq 500$ ml. Assume that we know true variance $\sigma^2 = 4$ ml². According to the critical region, should we reject H_0?

We want $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} < -z_{0.05}$ to reject H_0. The critical region is then $\left(-\infty, \mu_0 - z_{0.05}\,\frac{\sigma}{\sqrt{n}}\right) = (-\infty, 499.264)$. Since our data $\overline{X} = 499.115$ lies in the critical region, we reject H_0.

In[112]:= `z_α_ := InverseCDF[NormalDistribution[], 1 - α];`

$$500 - \frac{2\,z_{0.05}}{\sqrt{20}}$$
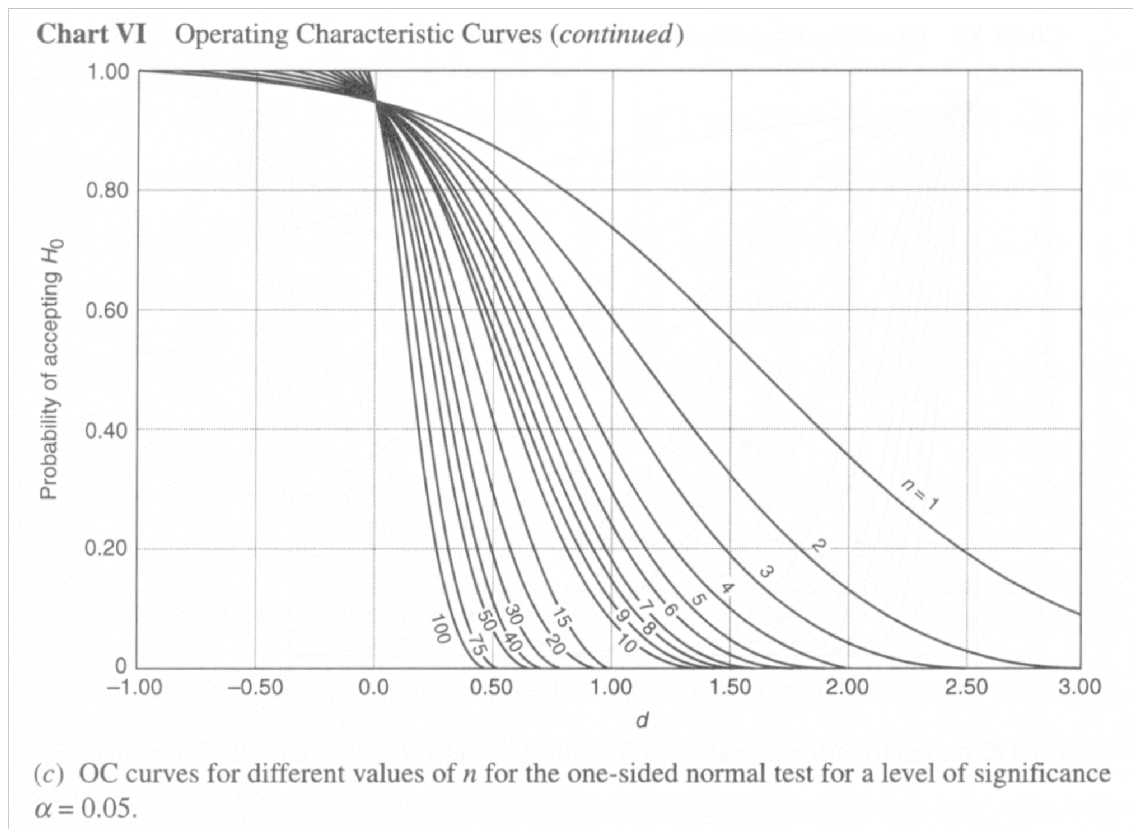
Out[113]= `499.264`

β and Sample Size

Definition: $\beta = P[\text{Type II error}] = P[\text{Sample statistics lies outside critical region} \mid H_1 \text{ true}]$.

Usage:

- To calculate a good sample size. $n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \, \sigma^2}{\delta^2}$.

- To calculate the **_power_** of a test. power $= 1 - \beta$. A powerful test needs smaller sample size, and is more likely to reject H_0.

- To plot the OC curve, or β vs. d curve, so that we find n more easily. Here

$$
d = \begin{cases}
\frac{|\mu - \mu_0|}{\sigma} & \text{if } H_0 : \mu = \mu_0 \\
\frac{\mu - \mu_0}{\sigma} & \text{if } H_0 : \mu \leq \mu_0 \\
-\frac{\mu - \mu_0}{\sigma} & \text{if } H_0 : \mu \geq \mu_0
\end{cases}
$$

Chart VI Operating Characteristic Curves (*continued*)

(*c*) OC curves for different values of n for the one-sided normal test for a level of significance $\alpha = 0.05$.

Examples:

What is the power of my test given that the true mean $\mu = 499$ ml, true variance $\sigma^2 = 4$ ml^2 and the sample size $n = 20$?

We have $d = \frac{|499 - 500|}{2} = 0.5$, $n = 20$. According to the OC Curve, $\beta \approx 0.3$. So the power is approximately 0.7. So the test is not powerful enough.

If the true mean of my product has volume $\mu = 499$ ml, and the variance is $\sigma^2 = 4$ ml^2, what sample size do we need to get a power of 0.8, if we set $\alpha = 0.05$?

We have $d = \frac{|499-500|}{2} = 0.5$. We will need about 30 samples to get a power of 0.8.

In[221]:= $\dfrac{(z_{0.05} + z_{0.2})^2\ 4}{(499 - 500)^2}$

Out[221]= 24.7302

Plot the OC Curve of the test of my product, given $\alpha = 0.05$, sample size $n = 20$ and the variance $\sigma^2 = 4$ ml^2.

In my case $H_0 : \mu \geq 500$, $H_1 : \mu \leq 499$, we need to calculate

$$P[\text{fail to reject } H_0] = P[\text{Sample statistics lies outside critical region}]$$
$$= P\left[\overline{X} > 499.264 \,\Big|\, \tfrac{\mu_0 - \mu}{\sigma} = d\right]$$
$$\leq P\left[\overline{X} > 499.264 \,\Big|\, \mu = 500 - 2\,d\right]$$
$$= P\left[\frac{\overline{X} - (500 - 2\,d)}{2\big/\sqrt{20}} > \frac{499.264 - (500 - 2\,d)}{2\big/\sqrt{20}}\right]$$
$$= P\left[Z > \frac{499.264 - (500 - 2\,d)}{2\big/\sqrt{20}}\right]$$

In[120]:= **Plot$\left[$1 - CDF$\left[$NormalDistribution[], $\dfrac{\textbf{499.264 - (500 - 2 d)}}{\textbf{2}\big/\sqrt{\textbf{20}}}\right]$,**

{d, -.5, 2}, AxesLabel → {"d", "P[fail to reject H₀]"},
PlotLabel → "OC Curve of Z test when n=20"$\Big]$

Out[120]=
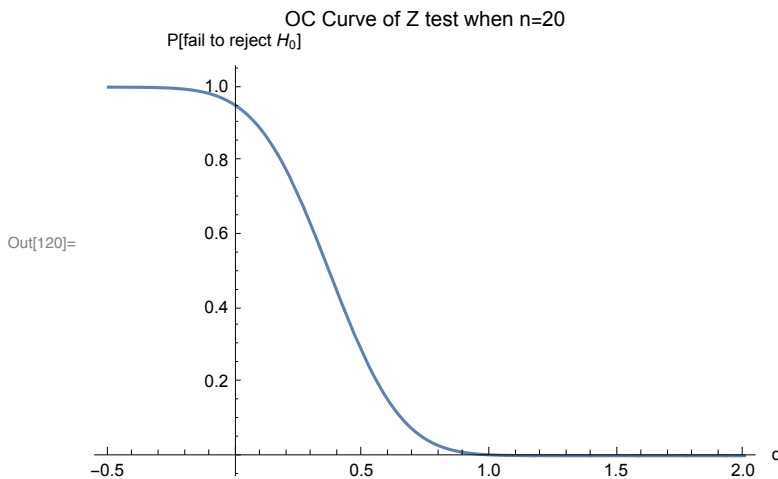
When $d = 0$, we can see that $\beta = 1 - \alpha = 0.95$.

How large should the true mean μ differ from $\mu_0 = 500$ ml so that H_0 is rejected 90 % of the time (given H_1 is true)?

According to the OC Curve, at $d \approx 0.65$ we obtain $\beta = 0.1$ or a power of 0.9. This means we need

$\frac{\mu_0 - \mu}{\sigma} = \frac{500 - \mu}{2} > 0.65 \Rightarrow \mu < 498.7$.

Null Hypothesis Significance Tests (NHST)

Hypotheses and Testing

Steps:

- Figure out the parameter θ and the null value θ_0 you are testing.

- We set up two competing hypotheses H_0 and H_1. (e.g. $H_0 : \mu \geq 500$, $H_1 : \mu \leq 499$). They can be one of the following: for some $\delta > 0$,

  - $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$.

  - $H_0 : \theta \leq \theta_0$, $H_1 : \theta > \theta_0$.

  - $H_0 : \theta \geq \theta_0$, $H_1 : \theta < \theta_0$.

- Decide α.

- Analyze data. Calculate the P-value and see if there is enough evidence to reject H_0.

  Note: NHST is biased against failing to reject H_0 (too easy to reject H_0).

Single Sample Tests and Comparison of Parameter

Test Statistics

From now on, you will learn a lot of tests. For each test, you need to pay attention to:

- What parameter are we testing?

- What parameter do we know?

- What is the distribution of test statistics?

Denoting the distribution of test statistics as $\mathcal{D}$, and d_α as the number such that $\int_{-\infty}^{d_\alpha} f_{\mathcal{D}}(x) \, dx = 1 - \alpha$, we reject at significance level α,

- $H_0 : \theta = \theta_0$ if $\mathcal{D} > d_{\alpha/2}$ or $\mathcal{D} < d_{1-\alpha/2}$. (Test statistics is too small or too big)

- $H_0 : \theta \leq \theta_0$ if $\mathcal{D} > d_\alpha$. (Test statistics is too big)

- $H_0 : \theta \geq \theta_0$ if $\mathcal{D} < d_{1-\alpha}$. (Test statistics is too small)

Single Sample Tests for Mean (Z-Test and T-Test)

Testing Parameter	Distribution of X_i	Variance σ	Sample Size n	Test Statistics	OC Curve x−axis (for two sided test)
μ	normal	known	any	$Z = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$d = \dfrac{|\mu-\mu_0|}{\sigma}$
	any	known	large		
	any	unknown	large	$Z = \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$d = \dfrac{|\mu-\mu_0|}{S}$
	normal	unknown	small	$T_{n-1} = \dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$	

Chart VI Operating Characteristic Curves (*continued*)
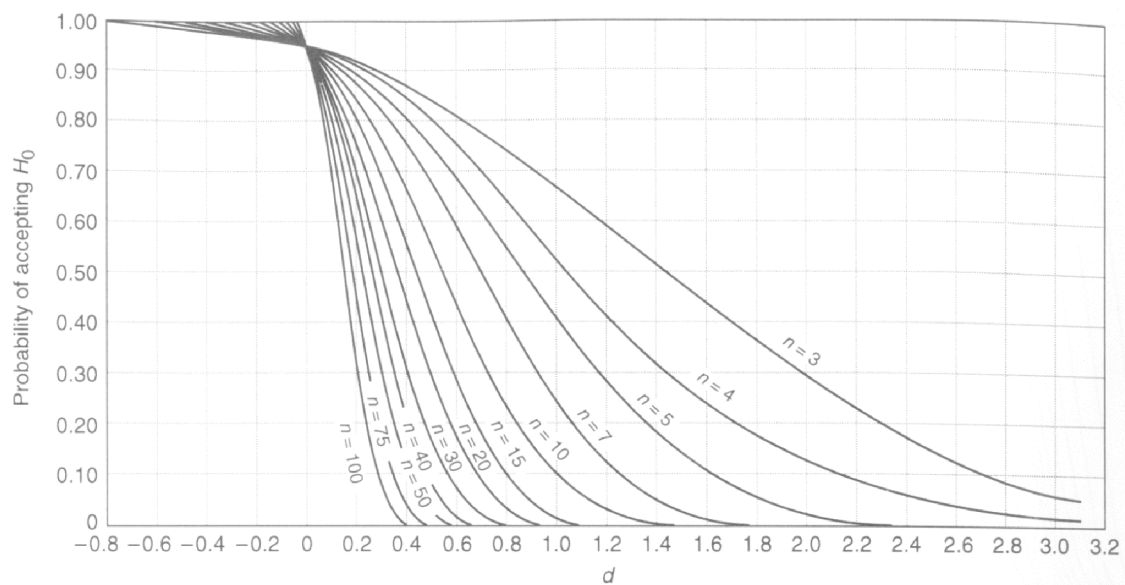
(*g*) OC curves for different values of n for the one-sided t test for a level of significance $\alpha = 0.05$.

Example:

If the true variance is unknown, what is the power of my T-test if the true mean is $\mu = 499$ ml?

In[127]:= $\dfrac{500 - 499}{\sqrt{\textbf{Variance[Data]}}}$

Out[127]= 0.462933

The corresponding power is about $1 - 0.35 = 0.65$.

Single Sample Tests for Variance (Chi-Squared Test)

Testing Parameter	Distribution of X_i	Test Statistics	OC Curve x−axis (for two sided test)

σ	Normal	$\chi^2_{n-1} = \frac{(n-1)\,S^2}{\sigma_0^2}$	$\lambda = \frac{\sigma}{\sigma_0}$

Single Sample Tests for Median (Non-Parametric)

Sign Test for the Median

Testing Parameter	Distribution of X_i	Test Statistics
M	any	$\text{Binomial}_{n,0.5} = Q_+$

where Q_+ is the number of data that lies above M_0. (Usually the data equal to M_0 is discarded).

Example:

In[128]:= **Data**

Out[128]= {499.728, 497.006, 498.871, 500.931, 500.157, 501.139,
495.516, 501.89, 496.894, 497.132, 503.624, 500.357, 498.011,
498.447, 496.074, 502.177, 498.081, 499.719, 497.657, 498.895}

Test the hypothesis $H_0 : M \geq 500$ ml.

We can count that $Q_+ = 7$. $P[Q_+ \leq 7] = \frac{1}{2^{20}} \sum_{i=0}^{7} \binom{20}{i} = 0.13$. There is no enough evidence to reject H_0.

Wilcoxon Signed Rank Test

Testing Parameter	Distribution of X_i	Test Statistics
M	symmetric	W_+

where W_+ is the sum of the positive ranks when ranking $|X_i - M_0|$. When there is a tie, we take the average value. For example, if the first, second and third smallest $|X_i - M_0|$ is a group of 3 ties, we rank them all as $\frac{1+3}{2} = 2$.

For small n, you can refer to a table of critical values here: http://users.stat.ufl.edu/~winner/tables/wilcox_signrank.pdf. For $n \geq 10$, we can approximate the test statistics with $E[W] = \frac{n(n+1)}{4}$, $\text{Var}[W] = \frac{n(n+1)\,(2\,n+1)}{24}$. For each group of t ties, the variance is reduced by $\frac{t^3 - t}{48}$.

In[●]:= **Sort[Data – 500]**

Out[●]= {-4.48438, -3.92626, -3.10595, -2.9944, -2.86761, -2.34263, -1.98916,
-1.91898, -1.55252, -1.12854, -1.1048, -0.281418, -0.271902,
0.156908, 0.356726, 0.930833, 1.13896, 1.89039, 2.17724, 3.62415}

Test the hypothesis $H_0 : M \geq 500$ ml.

We can count that $W_+ = 1 + 4 + 5 + 8 + 10 + 13 + 18 = 59$. According to the table, the critical value for $\alpha = 0.05$ is 60. Since $59 < 60$ we are able to reject H_0 with $P < 0.05$.

```
In[158]:= Needs["HypothesisTesting`"]
         SignedRankTest[Data, 500, "TestDataTable", AlternativeHypothesis → "Less"]
```

Out[159]=

	Statistic	P-Value
Signed-Rank	59.	0.0446939

Estimation of Proportion

For a Bernoulli random variable X_i, we can calculate $\hat{p} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$. The confidence interval for p would be

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1 - \hat{p}) / n}$$

If we want to test that $|p - \hat{p}| < d$ for some $d > 0$ with $100(1 - \alpha)\%$ confidence, we need the sample size to be at least

$$n = \frac{z_{\alpha/2}^2 \, \hat{p}(1 - \hat{p})}{d^2} \leq \frac{z_{\alpha/2}^2}{4 \, d^2}$$

Example:

```
In[●]:= Data
```

```
Out[●]= {499.728, 497.006, 498.871, 500.931, 500.157, 501.139,
        495.516, 501.89, 496.894, 497.132, 503.624, 500.357, 498.011,
        498.447, 496.074, 502.177, 498.081, 499.719, 497.657, 498.895}
```

Give a 95% confidence interval of the proportion of good product (with volume ≥ 500 ml) in my batch.

We have $\hat{p} = 7/20 = 0.35$ and $n = 20$. A confidence interval would be 0.35 ± 0.21.

Single Sample Test of Proportion (*Z*-Test)

According to central limit theorem, $\hat{p}$ is approximately normally distributed with mean p and variance $p(1 - p) / n$.

Testing Parameter	Distribution of X_i	Sample Size	Test Statistics
p	Bernoulli	large	$Z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$

Comparing Two Proportions (*Z*-Test, Pooled Test)

Testing Parameter	Distribution of X_i	Sample Size	Hypothesis H_0	Test Statistics

$p_1 - p_2$	Bernoulli	large	$\|p_1 - p_2\| \geq d$ for some $d > 0$	$Z = \dfrac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)_0}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$
			$p_1 = p_2$	$Z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ where $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

F Distribution

**Purpose**: To describe the distribution of the ratio of two scaled chi-squared random variables.

Mathematical representation: $F_{\gamma_1, \gamma_2} = \dfrac{\chi^2_{\gamma_1}/\gamma_1}{\chi^2_{\gamma_2}/\gamma_2}$ is said to follow an F-distribution with γ_1 and γ_2 degrees of freedom.

Properties:

- $P\left[F_{\gamma_1, \gamma_2} < x\right] = 1 - P\left[F_{\gamma_2, \gamma_1} < \dfrac{1}{x}\right]$.

- $f_{1-\alpha, \gamma_1, \gamma_2} = \dfrac{1}{f_{\alpha, \gamma_2, \gamma_1}}$.
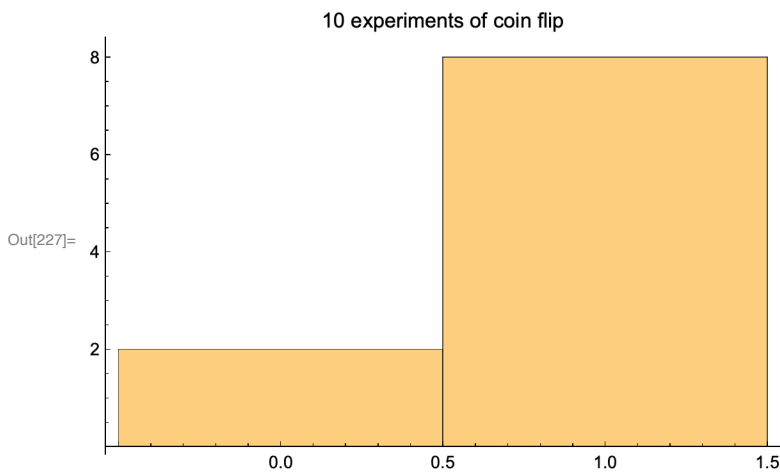
Comparing Two Variance (F-Test)

Testing Parameter	Distribution of X_i	Test Statistics	Hypothesis H_0	OC Curve x−axis $\left(\text{for } n_1 = n_2 = n\right)$
$\dfrac{\sigma_1}{\sigma_2}$	Normal	$F_{n_1-1, n_2-1} = \dfrac{S_1^2}{S_2^2}$	$\dfrac{\sigma_1}{\sigma_2} \geq 1$ $\dfrac{\sigma_1}{\sigma_2} \leq 1$ $\dfrac{\sigma_1}{\sigma_2} = 1$	$\lambda = \dfrac{\sigma_1}{\sigma_2}$
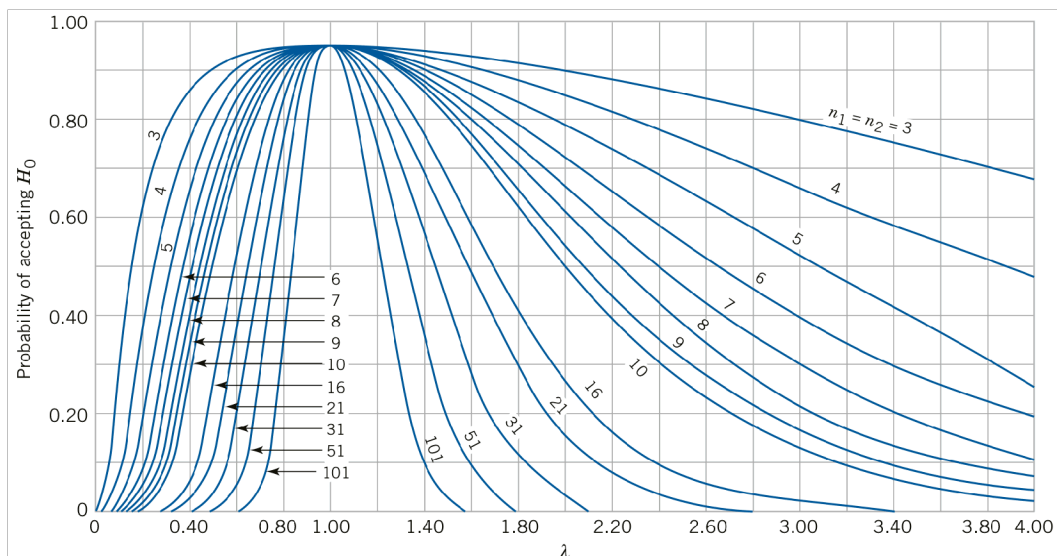
Problem For Discussion

Test for Fairness of Coin

In[227]:= `Histogram[RandomInteger[{0, 1}, 10], PlotLabel → "10 experiments of coin flip"]`

Out[227]=

In this discussion, we will explore the essentials for a test. To do this, we ask: How to test the fairness of a coin? The following questions arise.

- How to define the "fairness" of a coin? Does $p := P[\text{Heads}]$ have to be 0.5? Is $p = 0.52$ acceptable?

- How many experiments will you do for the test?

- If you get a certain result (say 8 heads and 2 tails in 10 trials), will you doubt the fairness of the coin?

- What is the probability that our conclusion is wrong?

- ...

We try to give an answer to these questions one by one.

Definition of "Fairness"

In the real life, p is likely to be normally distributed. In this case $P[p = 0.5] = 0$, meaning that almost all coin are not fair. To make it easier to calculate $P[\text{fairness}]$, we can define an interval for p, say $0.45 < p < 0.55$, such that if p lies in this interval, we consider the coin to be fair.

Frequentist Approach

We may calculate the probability of getting the data if $p = 0.5$.

$$P[\text{getting this or more extreme data}] = P[\,|\hat{p} - 0.5\,| \geq 0.3\,|\,p = 0.5]$$

$$= \frac{2}{2^{10}} \Sigma_{i=0}^{2} \binom{10}{i}$$

$$= 0.11 > 0.05$$

So there is little evidence that the coin is unfair.

Another Frequentist Approach

We may also think that the sample size is too small to tell us anything. To calculate the appropriate sample size, we take

$$n = \frac{z_{\alpha/2}^2}{4\,d^2} = \frac{1.96^2}{4\,(0.05)^2} \approx 385$$

samples to detect a difference of 0.05 and get a test of significance $\alpha = 0.05$.

In[191]:= `Mean[RandomInteger[{0, 1}, 385]] // N`

Out[191]= `0.496104`

We then calculate $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.496 - 0.5}{\sqrt{0.5^2/385}} = -0.156$. This gives a P-value of 0.876. There is no

evidence that the coin is unfair.

Bayesian Approach

We try to obtain $f_P(p) := P[P[\text{Heads}] = p]$. For simplicity, we use the uniform distribution $f_P(p) = 1$ for $0 \leq p \leq 1$ (In the reality the distribution may be more concentrated at $p = 0.5$). Using Bayes' theorem,

$$f_{P\,|\,h\,\text{Heads and }t\,\text{Tails}}(p) = \frac{P[h\,\text{heads and }t\,\text{tails}|P[\text{Heads}]=p]\,f_P(p)}{\int_0^1 P[h\,\text{heads and }t\,\text{tails}|P[\text{Heads}]=s]\,f_P(s)\,ds}$$

$$= \frac{\binom{h+t}{h}p^h(1-p)^t}{\int_0^1 \binom{h+t}{h}s^h(1-s)^t\,ds}$$

$$= \frac{p^h(1-p)^t}{\int_0^1 s^h(1-s)^t\,ds}$$

This appears to be a ***beta distribution***, and simplifies to

$$f_{P\,|\,h\,\text{Heads and }t\,\text{Tails}}(p) = \frac{(h+t+1)!}{h!\,t!}\,p^h(1-p)^t$$

Plug in $h = 8$, $t = 2$ gives us the following.

Out[220]=

Probability Density of P[P[Heads]=p]
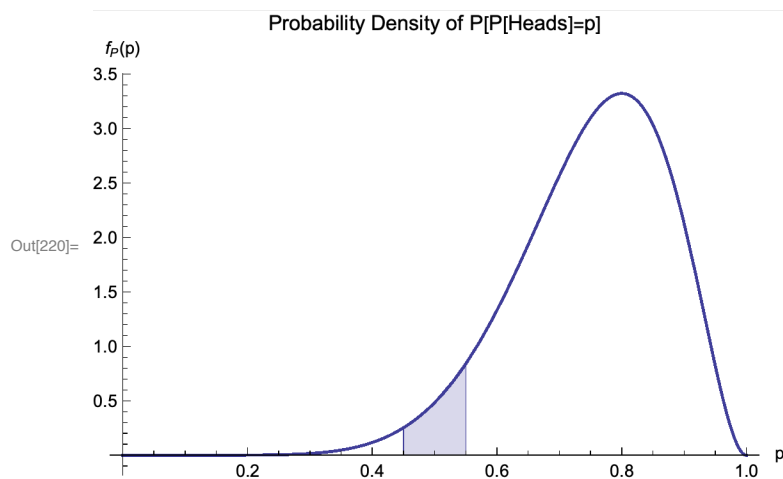
Then

$$P[\text{The coin is fair}] = \int_{0.45}^{0.55} \frac{(h+t+1)!}{h!\,t!}\, p^h (1-p)^t \, dp \; = 0.0504 > 0.05$$

The evidence is not strong enough to prove that the coin is unfair.